Data, Dependence, Democracy:
Influence in the Secondary Use of Government Information

By

Duncan S. Friend

Submitted to the graduate degree program in Public Affairs and Administration and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Chair: Charles R. Epp

Jacob Fowles

Marilu Goodyear

Rosemary O'Leary

William Staples

Date Defended: April 6, 2020

The dissertation committee for Duncan S. Friend certifies that this is the approved version of the following dissertation:

Data, Dependence, Democracy: Influence in the Secondary Use of Government Information

---

Chair: Charles R. Epp

Date Approved: May 8, 2020

Abstract

This study examines private sector use of information generated from government operations by focusing on how either or both parties come to mutually depend on such “secondary use” and the resulting impacts on government information, policy, and practice. Viewing government information as a potentially valuable economic resource, the study uses a theoretical framework that predicts that choices in the design of policies may result in unanticipated side effects that create or change incentives for interest groups, mass publics, or government decision-makers, resulting in actions that shape the environment in which the policy is implemented, creating dependencies, providing motivations or disincentives for action, and determining the bounds of the playing field for future policy decisions. To understand these relationships, the study focuses on cases in which one party is dependent and the other is not along with cases where there is mutual dependence. Using qualitative methods, including analysis of 250 documents and 65 interviews with the public and private sector employees involved, it investigates the extent to which the private sector seeks to, and gains, influence on the content, structure, or availability of government data and related policy and practice, how they attempt to exercise such influence, and its effects on the form or nature of the data, access policies and methods, decisions based on the data, or other government practices. The potential mechanisms and impacts of this dependence on secondary use of information are important to the study of government as lack of effective policy and controls to identify and manage its effects may allow the interests in or benefits obtained from this dependence to undermine the effectiveness of government programs, or weaken or divert government from its mission by affecting the nature of the data it uses, its priorities, resource allocation, or facets of its operations in service of these interests or benefits. The principal finding of this research is that such relationships exist and can have effects on

practice, including priorities and policy, and to some degree data or its format, as well as introduce market values into public decision-making, impacts that are largely unregulated by government policy.

Acknowledgments

The thing I will always remember most about this journey came at the end. In calls with friends and family about my successful defense, some of whom I had not talked with in a while, at first words of congratulations and memories, and then the conversation turned to their lives, and what was happening in their world. All these stories, all these dear people. The richness of those short hours was almost unbearable, in happiness, in nostalgia, in love.

This dissertation is really just one story of many and I feel lucky for the experiences that allowed me to tell it. More than most, I think, I have depended on the kindness of those around me, in school and the world outside it, often carrying me in ways large and small to this point. While I am, perhaps, a small part of your story, to me, each of your contributions has been enormous. The voices you hear throughout these case studies are from people who donated their time and perspective to help me understand things the way they do. I owe them all a great debt for their candor and willingness to engage a student in discussion about their daily work. Then, all along the way, was my advisor, Chuck Epp, helping me make sense of the stories and urging me out of the detail into a broader perspective on their meaning. This couldn't have happened without his kindness, patience, thoughtfulness…and resolute support. There are others, too, in the School to whom I owe a great deal. Marilu Goodyear, who, by asking me to return to help her teach a class one spring, really brought me into the program and then helped me gain my bearings, always in touch throughout. And, of course, her and the other members of my dissertation committee who contributed great ideas that improved my work and pointed me toward new directions for future research. I can't thank you enough both for serving on my committee and encouraging me to further pursue this area of study.

And then there's Diana Koslowsky. For her longtime advice, and commiseration, and her indefatigable efforts on my behalf, I wish to express my heartfelt thanks. She is the consummate administrator, and, to me, after all, that's what it's all about.

Any recounting of my journey would be remiss without an acknowledgment of the love and support from my friends. Like the people I've mentioned above, my praise and thanks fall far short of what they have done for me. There are too many to name here, yet I must include Floyd Rockers, who has never been far away throughout my time in the program and many years leading up to it, helping my wife and me in fair weather and storm (literally hooking electrical service back up to our house in an ice storm). Your constant encouragement has meant a lot and helped me get by with working and focusing on school at the same time. And Roberta Stadler, who, while dedicating her life to the church, has also dedicated it to her students, to which she has added me in an honorary capacity, both at church and in school. Thank you for all your support and checking up on me during the final stretch.

Then, there is my family. My mother-in-law, Betty Lou Wright, the light that will never stop shining for me. She accompanied me throughout this journey until fall 2018, always urging me onward as we talked almost every day. I am certain she is up there smiling. And, my grandmother, Bernice Douglas who raised me and her sister Eva, both of whom I know would feel the same way. In dedication, I end with the person who means the most to me, and is so much of who I am, my wife Elizabeth. There could never be a better example of endurance, and faith, and hope, and love than she showed me during this time. And yet, as she would tell me, it was my mom, Helen Swenson, who was right there all along beside me, too, in little ways that even she could see. So, then, to the two of them.

In the end, what we share is our faith in God and the benefit of His love and mercy. As Lula Collins sang "This Joy that I have, the world didn't give it to me." All thanks to Him.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

In 2014, the State of Texas received over $63 million in revenue from sale of data about drivers in the state (Texas Comptroller of Public Accounts, 2015). From 2010 through 2015, the Commonwealth of Pennsylvania received $157 million in revenue from similar sales (Blacher, 2016), and in the five year period 2005-2010, the company Choicepoint, Inc. alone paid the Ohio Bureau of Motor Vehicles more than $36 million for data about drivers in that state (Guillen, 2010). In turn, in 2018, a national company that assists state governments and industry in facilitating these transactions booked over $100 million in revenue for that service (NIC, 2018) on top of the charges paid by industry. The business model of the companies buying the data, primarily insurers, is dependent on the use of it for rate setting, with additional purchases made by information brokers who resell it to other parties for purposes allowed under state and federal law.

Private sector use of government data does not always involve revenue to government, however. In 2009, the federal government spent about $3.4 billion on meteorological research and operations that served as key data in driving the $1.7 billion private weather forecasting industry (U.S. Department of Commerce, 2014). They, and others, receive this data free of charge and use it to produce products and service. Examples include field-level hail forecasts used by farmers, or predictions for national railroads about locations where the heat and sun may result in bent track, a leading cause of derailments (Global Science & Technology, Inc., 2018b). In another program area, both state and federal archives that lack sufficient funding to perform the task have struck agreements with Ancestry.com (Ancestry, 2009) to digitize records of interest to genealogists at no cost in return for exclusive rights to sell access to the images to the public. And, there are companies without any agreements at all, whose primary business model is to aggregate and process data about restaurant inspections obtained free of charge from state and local government websites to resell to other companies (like insurers or franchise owners) interested in monitoring risk and code compliance.

In all these examples, we see industry business models dependent for revenue on some form of data created or held by government. In each case, industry benefits. In some cases, it appears that government benefits greatly as well - in others, perhaps not at all. It is important to note, however, that this data is collected or generated for primary use in the execution of a government program, established and governed by law, regulation, and policy. The private sector uses of the data, while they may serve a commercial need, accomplish some social good, or even contribute to the achievement of government's mission, are "secondary" (re)uses of the data, often with limited formal guidance or governance in their execution.

Although one might assume that these government programs would continue to collect and use the data in the same way whether or not the private sector becomes interested in reuse, the premise of this dissertation is that secondary reuse of government data can create a potentially powerful *mutual dependence* between government and private users of government data. The idea of such mutual dependence can be quickly illustrated by envisioning the extreme. Imagine that tomorrow, access to government data has been unleashed so that almost every conceivable public record is online, a paradise of transparency. This includes raw data from tens of thousands of computer systems containing non-personal information that most would agree would be non-controversial to distribute. In turn, again, tens of thousands of businesses come to depend on and monetize this data as part of new innovative business models, and, in another extreme assumption, government produces revenue from its sale to these private sector companies to partially support the government programs and computer systems that create the data (admittedly selling data is an option presently prohibited by policy at the federal level), or even other un-related government programs. And, each of these businesses using the data supported a further ecosystem made up of its paying customers, of course (how could they make a profit otherwise?).

Now for the rub. Suppose now that you, as a public manager, wished to replace a system used to run your program to deliver benefits to the public, or monitor streetlights, and the changes you propose would result in the collection of more or less data, or cause its format or frequency of production to change. Or the legislature abolished the program in next year's budget. With the change, the populations dependent on your program are not only the recipients of services, regulated parties, or public safety purposes it was designed for; they now include the secondary users of the data—the businesses that have grown up using the data as an essential part of their business model. This giant ecosystem of money-making business owners and their customers, many of whom may use the data for an entirely different purpose than that for which it is used in the government program that creates it - say, a commercial product that rates the value or safety of a neighborhood, or a service that predicts automobile sales - is still there, but suddenly without the data on which their business model depends. Would they not have something to say about your changes, the impact (including cost) on their business of having to accommodate your changes to the data, or the impact on them of shutting your program down? And what about the revenue or other benefits your agency gets from that business that could now be reduced, or go away completely? It is here, in these relationships, that my research begins.

The topic is timely. Interest in and advocacy for the secondary use of government information has been increasing dramatically, as has its availability in formats that lend themselves to easy reuse by computer programs. Open government and transparency-focused groups are actively lobbying, researching, publishing, and advocating in government and academic communities for policies to make governmentally-held data widely available to secondary users to increase transparency, accountability and good government – as well as to foster economic development. Private sector interests are also championing the potential economic benefit of open access to these data. In one prominent example, a study by Deloitte Analytics (2013) for the British government estimated the overall annual value of open data to all sectors to be as high as 6-7 billion pounds. In a survey of "public sector information" (PSI) reuse, Vickery (2011) estimates that "The aggregate direct and indirect economic impacts from PSI applications and use across the whole EU27 economy are estimated to be of the order of EUR 140 billion annually." This advocacy is producing results. At the federal level, in 2013 President Obama signed Executive Order 13,642, *Making Open and Machine Readable the New Default for Government Information* (EOP, 2013). State and local governments have followed suit, increasingly deploying open data portals and policies modeled along the lines of the federal open data portal, Data.gov (Sunlight Foundation, 2016) . New York University's Governance Lab project published the "Open Data 500" (http://www.opendata500.com/us/), a list of companies whose business models were driven in full or part by public data, mostly sourced from government. As another example, the Center for Open Data Enterprise (http://opendataenterprise.org/), a 501(c)3 nonprofit, has facilitated over twenty "roundtable" meetings over the last several years between federal agencies and business to examine opportunities for mutual benefit from government data as part of its mission to "maximize the value of open government data for the public good" (CODE, 2020).

Problem Statement

Over time, some private sector companies have come to depend on the use of government data to produce revenue. In turn, some government agencies that produce this data have also come to depend on revenue from its sale to these companies, or for other benefits from this private sector use. When data serves as an economic resource, or even a subsidy to industry, the theories to be discussed below suggest that the relationship may take on characteristics similar to other cases where the government provides benefits, leading these parties to attempt to influence government policies and practices to ensure a continuation of these benefits, as well as to maintain or increase their value. If this occurs with government information, the impact of these influences would be significant because of the vital role information - and access to it - plays in the operation of government, and because choices about the design, creation/collection, retention, and access to data are subject to significant discretion on the part of public administrators. The danger here is the potential for a form of institutional corruption, in the sense defined by Lessig (2013):

> Institutional corruption is manifest when there is a systemic and strategic influence which is legal, or even currently ethical, that undermines the institution's effectiveness by diverting it from its purpose or weakening its ability to achieve its purpose, including, to the extent relevant to its purpose, weakening either the public's trust in that institution or the institution's inherent trustworthiness.

Research Questions

The central question of this dissertation is: How does the mutual dependence between industry and government on the secondary use of governmental data affect decisions regarding policies and practices for data access, or the composition of the data, or even priorities in the related programs themselves in the interest of maximizing benefits to one or both parties?

My research assumes that policies and practices related to access to governmentally held data have resulted in private parties developing an interest in access to those data and in shaping the policies on access to that data, if not the data itself, to better serve their interests. My general question may be divided among several particular questions:

1) What private parties have developed an interest in governmentally held data?

2) To what extent do these parties seek to, and gain, influence on the content, structure, or availability of these data?

3) How do they exercise this influence?

4) What are the effects of this influence on the form or nature of governmentally-held data, policies on data availability, and decision-making based on these data?

The following sections review the theories that underlie these premises and questions, and propose a set of expectations derived from their application to this topic.

Central Concepts

Information as a Resource

One key concept in this relation of dependence is that government information is a resource, a resource with potential value in an information market. The government controls vast quantities of information, ranging widely across almost every conceivable topic: the demographic characteristics of the population, their driving behavior, their spending habits, business practices and employment patterns, agricultural productivity, the extraction of minerals and fuels from the earth, the condition of buildings and roads, the weather and climate—and it goes on. Public policies (The federal Freedom of Information Act, or state and local open records policies) require much of this information to be available to any member of the public (or business) upon request. However, a convergence of powerful improvements in technology for the electronic collection and processing of information in both the public and private sector, along with high-speed and high capacity networking have expanded the scope of what these policies are being used for beyond transparency and accountability. When viewed through this lens, government data is a commodity used by business, a starting point in a supply chain that can lead to significant revenue. And, when used for that purpose, these transparency policies now perform the function of resource allocation, absent much of the detail and guidance one would expect of government for resources of such value. The focus of the requests and resulting responsiveness also moves from individuals and organizations concerned with accountability to major industries who rely on government data to produce revenue and - in government - agencies that rely on revenues from its sale to them, cost avoidance in accomplishing their mission, and/or building constituency by demonstrating the usefulness (and existence) of their programs. Yet, government managers exercise wide latitude in what is made available online vs. upon request, and the format and technologies used to provide it.

Information Construction

Another concept key to the consideration of influence and its impacts is that, while it is common to assume that information is “objective,” the premise of this dissertation is that information is always *constructed* in ways that shape its form, content and use. An additional premise is that this information-construction significantly affects how the information is used, to whom it is useful, and the relationships of exchange and power that grow up around it. And government information is a resource like no other. Information produced, collected, and disseminated by government results in a representation of its own internal operations, the world in which it operates, and its actions in society. Data received, created, and collected about the physical, social, and political environments are used in problem definition, policy formation, and performance measures. The resulting documentation serves as evidence of the appropriateness and legality of decisions and actions by public officials, the need for and effectiveness of programs, and forms the basis through which citizens express preferences and hold their government accountable. In turn, decisions made during the life cycle of information about what data to collect, what to keep and how long, how to process it, and what access to allow to it and by whom have consequences for critical aspects of society from identity, individual (and property) rights, to commerce, public health, and protection of natural resources, to national memory - and democracy itself.

Prior Research

Data Reuse by the Private Sector

Although there is some research directly on the reuse of government data by the private sector, in the main it is not very helpful for this study as it does not address the question of mutual dependence. The academic research on reuse of public sector information to date has been mostly focused on its role in transparency and accountability and related impacts, along with studies of its success. Thus, scholars have examined the benefits of increasing the availability of governmentally-held data, along with pitfalls and inadequacies of increasing this availability (Janssen, Charalabidis, & Zuiderwijk, 2012; Davies & Bawa, 2012; and the potential inequity in access to and use of it (Gurstein, 2011; Johnson, 2014; Ruppert, 2015). Research has also addressed privacy concerns, especially to the extent that such information may be legitimately obtained by credit bureaus or insurance companies and used to profile and "score" citizens and businesses (Citron & Pasquale, 2014; U.S. Federal Trade Commission, 2014). Work has also begun to look more closely at the politics of reuse of government data. Catlaw & Sandberg (2014) explore how federal open data policies reinforce neoliberalism in the United States by putting the emphasis on the responsibility of individuals enabled by access to data. And, Bates (2014a) frames policy approaches on reuse in the United Kingdom as intended to advance neoliberal political agendas of privatization of public services and assets.

During the period that I have been conducting this research (2016-), a handful of studies have touched on its context and subjects. One paper, "Rawification and the Careful Generation of Open Government Data" (Denis & Goeta, 2017) used ethnographic techniques to identify that, far from external access to "raw" data, there was significant work occurring inside the bureaucracy to transform data to make it intelligible and usable for release – essentially making it "manufactured," a finding relevant to this study. A related phenomena is found in research by Heimstadt (2017), who sees organizational responses to transparency pressures addressed by "decoupling," essentially shaping the data to be provided externally through selective omission or creating new versions of it in a process he calls "openwashing." Quite recently, a qualitative analysis (Rujier, et. al, 2019) was published studying two cases of open data, one in France and the other in the Netherlands examining the "politics of open government data" and institutional responses to pressures to expand data access. However, their focus was on data mostly of public interest for accountability or societal benefit, vs. private use. The authors identified government's use of "strategically opaque transparency" to reveal information in some domains but not others, a subject they urged more research into. Unique here was their framing of their work on open data using theory about more general responses to institutional pressure. Another qualitative study, released in the last month (Tupasela, Snell, & Tarkkala, 2020) also looks at two European countries, Denmark and Finland, to identify challenges and complexity of the interests of the state and its citizens as these governments consider how what the authors (and government) refer to as the "Nordic gold mine" of personal data, especially related to health, could be used by the public and private sector. In its relevance here, they find in concept that the efforts produce "new and more complicated dependencies between the state, companies, and data sources" (Tupasela, Snell, & Tarkkala, 2020). This study is too new for me to incorporate fully into my research, but these more recent publications reinforce the idea that the research questions

I propose would be relevant and of interest to scholars involved and the academic research communities they represent. And, there is recent interest by a larger public in the specific case to which much of my study is devoted. Borowitz (2019) published a brief article in *Science* that made note of the trend toward cloud-hosting of government data that included reference to activities at NOAA and NASA. The focus of the article was a concern that these models might negatively affect public access and scientific research due to the costs involved and private sector's need to recoup them somehow – a concern that top of mind in the model that NOAA attempts to create in its Big Data Project.

Construction of Government Information

Although most studies on the reuse of governmental data do not address the issue of mutual dependence, a number of other studies address one central building block of this dissertation: the social construction of governmental data. These studies confirm that the content, structure, and/or availability of governmentally-held data are for the most part a result of choices made during the creation or collection of data. In "The Social Construction of Documentary Reality" (1974), sociologist Dorothy Smith lays out the circumstances, motivations, and processes associated with the creation of documentation by individuals in an organization. She refers to the process of collection and creation of the documentary record as the "production of accounts," noting that these artifacts "stand in for an actuality that is not directly accessible." (Smith, 1974). For Smith, decisions about what to record involve a recognition of normative expectations about what constitutes facts and their arrangement. Smith proposes that the data chosen anticipates potential audiences and their expectations, keeping in mind the need to present the finished product in a way that will be accepted as a "fact." This theory has two practical implications for my study: First, it suggests that choices about data anticipate user expectations. While this might seem obvious, it should be noted that such expectations might conflict, especially between "primary" users focused on program execution and "secondary" users focused on revenue production, which puts more focus on the priorities and incentives for those making the decision. Second, her theory emphasizes that those involved in the creation or collection process are well aware of the need to create the information in a way that masks this subjectivity if the resulting data is to be taken as fact.

Smith's theory takes on further relevance in analyzing external influence on the "shape" of information when coupled with the work of Feldman and March (1981). In "Information in Organizations as Symbol and Signal" the authors focus on explaining the seeming incongruity between how the rational model explains decision making and observations about how information is actually created and used – or not used - in organizations. One outcome of their research is the finding that information can serve as a "symbol of competence," reinforcing external perceptions of the rationality of organizational decision making and, in turn, the legitimacy of the organization. As an audience for the symbolic dimension of the collection, display, and use of information, then, external parties' perception of information quality (objective, rationally collected or created) are important. The symbolic role of information in maintaining legitimacy could serve as a mediating factor in the exercise of private party influence on information. If information is core to decision-making and "a good decision maker is one who makes decisions the way a good decision maker does" (Feldman & March, 1981), then private party influence on information will be constrained by the need of the organization to continue to appear legitimate in its collection/creation and use of the data. In turn, this implicit association of increased information collection with increased legitimacy may open the door to collecting additional information that reinforces this impression, but, in actuality, provides value mostly to secondary, rather than internal government users.

Trace (2002) surveys the work of a number of sociologists who have examined the processes of recordkeeping in law enforcement and draws lessons that can be helpful in considering influence in the construction process. She found that ethnographic studies identify three overlapping influences: social factors that influence whether or not information is recorded; a tension between the "use" (technical or instrumental) and "purpose" (symbolic or to promote a particular view of reality) of records; and, the impact of anticipating uses by internal and external audiences on decisions made about its form and content. In the first case, she uses the work of Morash (1984) who looks at the socialization of police work and how policemen are taught through the culture of policing to document particular things and to avoid documenting others in ways that deliver a representation of the situation that meshes with anticipated uses in the court system. In the second, she cites Cochran, et. al (1980) whose work is concerned with "proactive" recordkeeping, where those creating documentation have an understanding of its potential use and make a number of decisions prior to committing information to paper that incorporate assumptions that end up shaping the record (or, in Smith's case above, the "fact"). Van Maanen and Pentland (1994) refer to both the "technical" use of records by police to help in remembering actions and to aid decisions and to their "rhetorical use," where the account is consciously structured to create an impression, leading to "the institutionalization of a particular view of reality represented in part in documentary form by a highly specific and specialized form of language, order, and form" (Trace, 2002).

Consideration of these influences and the related decisions are key to identifying impacts of private parties on the process of constructing government information. The role of the individual(s) in determining what to record, and what not to, is important, along with the culture of the organization that frames the subjectivity exercised in creation. In this light, recordkeeping systems can also be conceived of as "forgetting" systems (Bowker, 2005), producing the official organizational account of activity and, consequently, establishing the facts that will not be considered important by omission. Van Maanen and Pentland's (1994) work considers the information produced by an organization as a "structural attribute" of it, performing a legitimizing function in addition to its instrumental role, one complementary to and reinforcing the impression of the values the organization wishes to convey (see previously Feldman & March, 1981). These studies also suggest, like Smith (1974), that *the anticipated internal and external uses for the data are incorporated in these decisions*. This role of information as a structural attribute and its interaction with the shape of data is also emphasized in the discussion of resulting "knowledge infrastructures" (Bowker, 1996; Edwards, 2010) that underlie the creation and management of scientific data. Edwards (2010) defines them as "[R]obust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds." For example, the development of weather models must take into account the fact that "natural" data are not collected for every point in the environment (vertically in the atmosphere or horizontally across space), so factors like individual judgment, capabilities in technology, and professional norms come into play in making decisions to fill the gaps that allow those models to "work" (Edwards, 2010).

Similar research in Science and Technology Studies supports the idea that there are other dimensions of information construction that may be relevant in identifying both impacts of influence by private parties and in considering how resulting changes may affect decision making. As is the case with weather data, categorization and standards affect the way information is structured and, consequently, define what is and is not recorded and the range of available values data can take on (Bowker & Star, 2000) with implications for resulting decisions. Classification systems (Star, 1999) and standards may be freighted with meaning (Lampland & Star, 2009) that can influence how data is interpreted by internal and external users. By definition, they serve as "boundary objects" (Star & Griesemer, 1989; Bowker, 1996) that are often imposed or adopted from the external environment for the purpose of communication and data sharing.

The nature of the resulting influence is hard to predict and may depend on matters of timing. For example, the absence of clearly-established standards or data categories may increase the influence of secondary users on the data-construction. But having standards may not diminish the influence of secondary users. Powerful secondary-user interests may influence the construction of standards or data-categories at the outset; years later, while these may appear to be "just the way things are" in fact the interests of secondary users may be built into the structure of the data. By contrast, standards and data categories may be influenced mainly by professional or technical considerations that are entirely independent of the interests of secondary users. There is a third possibility, too: the interests of secondary users may influence these professional or technical considerations, which then influence the construction of the data. Whether or how any of these possibilities shapes the process is a matter for empirical analysis; but it is clear that attention should be paid to standards and data categories and influences on them.

Although studies of the social construction of governmental information are a key building block of this dissertation, these studies do not address the essential *policy context* of the construction of governmental information. For insight into how that policy context may shape this construction along with access to the resulting information, we must turn to another area of research, on what is called "policy feedback."

Policy Feedback

Private-sector reuse of governmental data virtually always makes use of data that are generated pursuant to formal policies. Understanding how these policies shape the data, and how these policies may become the object of efforts to influence the usability of the data, is essential. The starting place for gaining an understanding of these dynamics is Schattschneider's (1935) assertion that "new policies create a new politics." That is, choices in the design of policies may result in unanticipated side effects that create or change incentives for interest groups, mass publics, or government elites, resulting in actions that shape the environment in which the policy is implemented, creating dependencies, providing motivations or disincentives for action, and determining the bounds of the playing field for future policy decisions. While policies that restrict or enable access to information, such as the Freedom of Information Act (2012) or state open records laws are not "new," others, like federal open data policies (EOP, 2009; EOP, 2013) have been promulgated in the last decade. And, as noted earlier, both the public and private sector are now promoting greater access to and reuse of data in the name of economic development and social good.

The central tenets of this policy-shapes-politics theory were laid out by Pierson (1993), and built on by a series of scholars applying these concepts to policy choices in such diverse areas as civil war pensions (Skocpol, 1995), welfare (Hacker, 2002), social security (Campbell, 2003), taxation, higher education lending, and health care reform (Mettler, 2010; 2011). Pierson divides the types of policy effects that may generate feedback into two categories. The first are the effects of "resources and incentives," the former related to benefits provided by or resources created by the policy, and the latter those elements of policy design that create or change incentives in ways that influence or limit choices of those affected. The second is what he terms "interpretive" effects, that is, how policy design affects "the manner in which social actors make sense of their environment" (Pierson, 1993). In both cases, the impact of the effects is not limited to parties outside government, but may also impact individuals and resources inside it.

Resources and Incentives. According to Pierson, the resource effects occur when resources are provided directly to a population and, in turn, they become willing and able to act politically to defend that interest. A recent study of private parties engaged in government data reuse (Deloitte Analytics, 2013) identifies several business models benefiting from these information resources, including "aggregators" – organizations that collect and aggregate data; "developers" – individuals and firms that design and support software to facilitate data reuse; "enrichers" that use data to enhance their existing products; and "enablers" who facilitate the supply of data – a category into which "data brokers" who collect and resell data fall. The U.S. Department of Commerce recently estimated (U. S. Department of Commerce, 2014) the size of what it calls the "Government Data-Intensive Sector" of business in the United States. This sector, which includes "value-added re-packagers (of data), analysts, and bench markers," employs almost 90,000 people and produces roughly $24 billion in annual revenue. In one form or another, this suggests that there are enough resources in play to incentivize actions to maintain, if not enhance, access to governmentally-held data.

Intermediaries between government agencies and secondary users may play a particularly important role in these dynamics. Pierson theorizes that policies can help create "niches for political entrepreneurs, who may take advantage of these incentives to help "latent groups" overcome collective action problems" (Pierson, 1993). This theory can be applied to a consideration of how private parties may influence the content, structure, and availability of information. For example, in more than 25 states, a private company, the National Information

Consortium (NIC), serves as an intermediary that advocates for and produces revenue from the resale of government information to interested users (NIC, 2015). The roles, incentives, and activities of such intermediaries – to the degree that they exist in a particular market - along with their relationship to private data users are thus relevant to any study of private party influence on government information.

As policies are carried out, Pierson also proposes that they may impact the administrative capacity of the state. For example, policies may require the development or acquisition of specialized skills, or investment of resources in the infrastructure and relationships needed to support program delivery. Investments like these, both by government and those external parties impacted by policy, can create what Pierson (1993) calls "lock-in." These effects may result in path dependence by interest groups and other constituents, where the sunk costs make alternative approaches less attractive, even if feasible and more efficient. One important observation about lock-in is its tendency to "depoliticize" issues (Pierson, 1993), where the growing benefits accruing from the status quo lead to diminishing interest in or conflict over alternatives to it. Finally, Pierson also cites Arthur (1989) in introducing the idea of "adaptive expectations," where early participants in the policy ecosystem are in the position of placing bets about the future direction of policy and resources. These choices also become, in a manner, sunk costs, as the participants gain improvements from experience and coordination with other parties, potentially driving out the possibility of alternatives.

These concepts – lock-in, de-politicization of the activity, and path dependence from early policy choices – can easily be applied to the ecosystem of reuse of governmentally-held data. Government agencies invest to make data more easily accessible, both through technology and policy. In turn, both they and the private parties reusing the data are likely to learn together about how to make the relationship smoother, potentially tailoring aspects of it to accommodate particular methods and types of use that are adopted early on. For example, contracts or informal agreements may be created to ensure the ongoing benefit of such investment, and, as a result, there is little controversy or even visibility (see the related discussion of interpretive effects below) to other parties outside the industries that reuse the data – and thus few incentives to mobilize to restrict or eliminate these activities.

Interpretive Effects. While resource-related effects are central to understanding the way policies impact political arrangements, so, too, are the messages that are embedded in and conveyed by policies. Target populations for policies may include groups for which mass publics carry negative connotations, affecting the support for a policy, and how/if interest groups organize around it. Béland (2010) discusses how ideational and symbolic legacies of values and ideals can also be used to frame policies to influence outcomes, providing an example of how the heavy social and political connotations of the term "security" in Social Security were leveraged against a characterization of privatization legislation as a "gamble" to defeat it. As it relates to the reuse of governmentally-held data by private parties, these relationships appear to benefit from the symbolic legacy of freedom of information and transparency, as well as their close association with accountability and democracy, making efforts to mobilize mass publics against reuse more difficult. As a practical matter, then, one might see relatively little examination of the phenomena of private party reuse of governmentally-held data, or even influence on it, in the press or legislative activity that would challenge or restrain it.

Pierson (1993) emphasizes the role of "policy learning" as well, where the implementation experience of particular policies may be seen as positive or negative, affecting future choices about design of other policies. Such learning can also extend to recipients. From this, one would expect to see some similarity in policies regarding reuse of data by private parties across government organizations, and perhaps even a converging path toward a common approach to certain types of information, with diffusion driven either by positive or negative experiences – or by either government or the private party users. To the degree that the method of delivery of resources chosen in the policy design is "submerged" (Mettler, 2011), that is, relatively invisible to the public – say, money received by way of a tax credit vs. a welfare payment – it can affect how or if people and interest groups are mobilized around it (Pierson, 1993). This "traceability" of the benefits, or what Pierson, quoting Arnold (1992), refers to as "the length of the causal chain" can also be a factor in whether or not interest groups and the public connect outcomes to the impact of a policy. Pierson notes that these choices in the length of the chain are part of policy design, along with associating the impacts with specific decision makers (1993) and can be adjusted with political ends in mind. In considering the business models described above, the causal chain for the reuse of governmentally-held data can be quite long, occurring behind the scenes. Data obtained from government may be coupled with and perhaps indistinguishable from data obtained elsewhere then resold or repackaged for a variety of purposes. This, too, works to depoliticize this activity and may leave open more opportunity for exercising influence on the content, structure, or availability of the data without external scrutiny or mobilization of interests that might limit such influence.

Other Effects. In a review of the literature that has grown up around this theory, Béland (2010) draws attention to another effect not noted by Pierson, drawn from the research of Jacob Hacker (2002). The "influence of private practices on public policies" proposes that the design of a public policy may be influenced by existing or antecedent policies for similar or complementary programs in the private sector. These private sector programs can create the same sorts of effects, like lock-in, that public policies do, making it difficult to implement a public sector alternative like, for example, national health insurance when private health insurance has become an established private sector practice (Béland, 2010, citing Hacker, 2002). This observation is especially relevant for reuse of governmentally-held data. Essentially another aspect of path dependence, when applied to this area it suggests that the history of certain types of reuse of private sector data, say, by the insurance industry to determine risk, may establish precedents that can frame policies, expectations, and disposition toward facilitating reuse of data in the future, and even soliciting feedback and direction in structuring the relationship to support this reuse. While not exactly the same situation as anticipated in the example provide by Hacker, it nevertheless places emphasis on the impact of the precedence of private sector practices and the related expectations established in the acquisition of governmentally-held data.

Factors Constraining the Policy Feedback Process. Finally, some attention must be paid to the situations that serve to weaken the effects described above. Previously, the role of information in promoting organizational legitimacy has been suggested as a possible counterbalance to private sector influence. However, Patashnik and Zelizer (2009) explore how the effects of the policies themselves in producing influence can fail to occur, or diminish over time. The first, weak policy design, is fueled by the practical focus of most policy analysts who may not adequately consider how to encourage positive feedback effects. They also note that policies may be implemented for mostly symbolic reasons, without heavy focus on the success of implementation, citing cases like income tax policy before World War II, where little of the population was affected and little tax revenue produced. They also find, like Pierson (1993) that there must be enough resources in play to incentivize the formation and/or mobilization of interests. A policy may also be hamstrung by inadequate or even conflicting institutional supports. This depends on whether the government institutions affected by the policy have conflicting goals or cultures, or mutually support the policy (Patashnik & Zelizer, 2009). They may, for example, be embedded in networks or other relationships that reinforce a different, perhaps higher priority, set of interests that conflict with the desired policy outcomes.

These observations are directly relevant to policies that constrain or enable the reuse of governmentally-held data. In the first case, transparency-related open records or information policy may, in practice, be fairly easy to work around through classification of information, for example (Roberts, 2006), or even by failing to record information (Eriksson & Ostberg, 2009). Or, from a practical perspective, access policy may be focused on data in which private parties have little interest. Each of these cases, the latter for obvious reasons, would be likely to result in little effective influence. As for conflicts among institutional supports, goals, and cultures, research by Peled (2011) on the federal government's Open Data program references extensive literature to support his findings that information sharing is tied up with the calculus of winners and losers of power between organizations and within organizational factions that create and use it. In fact, data can become so valuable as to be used as "bargaining chips in agency trade" (Peled, 2011), in this case resulting in a significant constraint on participation in the federal government's Data.gov website that makes data available to the public and private parties. These findings suggest that where policy exists and might potentially incentivize influence by private parties who seek to reuse it, countervailing organizational power – especially in cases where the data may have value to other governmental units – may constrain this influence. On the other hand, it seems this might also open the door for influence by private parties who could align (and perhaps submerge) their interests with other agencies who also seek to obtain or change the data.

In considering mitigating effects, it must be understood that policies that restrict or enable the availability of data may also shape its construction, in the sense that data that are not available, or only partially available, can change the representation of the associated organizational or societal reality they are taken to reflect. Overman & Cahill (1990) propose that choices about information policy represent a compromise between "restrictive" values, such as usefulness, cost/benefit, secrecy, security, ownership, and the protection of privacy, and "distributive" values, like freedom of information and an individual's access to his or her own private information held by government. Policies on retention of data and access to it "bake in" compromises (such as a balance between private data held by government and open access to it) that have implications for the study of private influences on information access policies. As a practical matter, then, parties that seek to change well-established policies on retention or access (or create new ones) are likely to be constrained by this tension between information policy values and would likely gravitate toward changes that – at least on the surface – acknowledge and preserve them, rather than risk mobilizing forces against the change.

Expectations

To summarize my theoretical framework, government information is constructed. That is, it reflects choices by government about what data to collect or create (and what not to), how often to collect or create it and at what level of detail, how to organize it, how long to keep it, and which parts may be accessible to whom, and by what means. In turn, policy and practice concerning these decisions are likely to be shaped by the interests of private secondary users of the information and the interests of government in generating revenue or other benefits from the private sector use of this information - constrained by existing rules, norms and assumptions about access to information.

My principal expectation is that the process of information construction and distribution may be shaped by a process of *mutual dependency* between the government agencies that make information available and the organized interests that gain access to it. Many other things are likely to influence information construction and policy as well, but my focus will be on this process of mutual dependency. The basis for this expectation is a growing body of research showing that policies construct private interests, mobilize constituencies, and, over the long term, may fundamentally reshape the state-society relationship (Pierson, 1993; Skocpol, 1995; Hacker, 2002; Mettler, 2002; Campbell, 2003; Campbell, 2012; Mettler & SoRelle, 2014).

In cases where the government makes resources available, this body of research suggests that constituencies dependent on these resources are likely to attempt to influence government policies and practices to ensure a continuation of these benefits, as well as to maintain or increase their value. This is significant for several reasons. New influences may be at play in constructing government data, coming from parties with potentially different motivations than government or the public for determining its content and availability. Second, these parties, interested in the data as a resource, may be using it for different purposes than were intended in the government program. Finally, as a result of these potentially different motivations and uses, they may have incentives to restrict or expand its availability, change its content – in frequency of collection, quality, size, or otherwise - or take other actions related to the data, all of which have the potential to change government policy and practice in this area, or even decisions and practices that rely on the data.

These expectations may be specified more precisely as follows:

1) Private parties ("secondary users") that depend on government information are likely to try to influence the character and availability of the information in ways that favor their business

model.

a. First, and foremost, the secondary users will attempt to reduce uncertainty in the availability of the information, pressing for policies or developing other mechanisms like contracts that ensure its reliable quality and supply.

b. They are also likely to try to lower both the direct and indirect costs of obtaining the information. This may be through attempts to lower the fees charged to obtain it, but also by other changes, such as standardizing the formats and technologies used to supply it in ways that favor their reuse, and, perhaps, limit the usability to competitors.

c. Where there is an economic advantage to do so and the data is time-sensitive, secondary users may attempt to influence government to improve the timeliness of the data by increasing the frequency of its collection or creation. Or, they may press governments to provide notifications that alert them to changes, allowing quicker action and also reducing costs for the secondary user by helping them avoid paying to acquire redundant copies of (unchanged) information.

d. Where advantageous, secondary users are likely to attempt to increase the level of detail and breadth of the information collected and provided.

e. Where secondary users are not allowed to "stockpile" or maintain databases of information obtained, they are likely to attempt to influence government to retain the information they need for as long as required for their business purposes.

2) Government agencies that provide information to private secondary users are likely to take action to enhance revenue or other benefits from the sale or provision of this information.

a. Government agencies are likely to raise fees to deliver information to commercial third parties in excess of the actual cost to provide it. In turn, they are likely to put in place polices and agreements that allow third parties to depend on a consistent supply.

b. They are likely to customize the methods of delivery and the format in which data is provided to make its reuse by third parties easier. This can include improving the metadata (data about the data) and access methods available (such as real-time or bulk access).

c. Government organizations, within the constraints described below, may create new "information products" that enhance the value of the data. This could include increasing the frequency of collection or availability, the length of the period covered by the data, or expanding or otherwise enhancing the content of the data, or integrating it with other government information valuable to the secondary user.

d. Choices about data construction are likely to incorporate the decision makers' perceptions of user expectations. Thus, to the degree that decision makers are aware of private party re-use, their understanding of those parties' expectations is likely to be considered in the construction process. It is not clear, however, whether or not it will result in actions that affect the data.

e. Choices about data construction are likely to incorporate an awareness of the symbolic role of information in maintaining organizational legitimacy and will be made in ways that mask subjectivity and promote the legitimacy of the program producing it. This may serve as a mediating/mitigating factor in private party influence on information. That is, influence by private parties is likely to be confined to achieving impacts that are seen as legitimate by stakeholders and regulators, or perhaps result in some mutual reframing of the endeavor to increase its legitimacy in the public interest.

3) These interested actions by secondary users and government agencies are likely to be

conditioned on the ease and available methods with which the data, format, or policies on access and distribution – in whole, or in part – can be changed within the constraints of law and institutional norms.

a. In general, the less restricted and burdensome the modification, the more we may expect to see secondary users and government agencies act in the ways described above.
b. The more the agency may make "profitable" revenue or reap other benefits from sale or provision of the data, or anticipate doing so, the more we may expect to see the agency act in the ways described above.
c. The more that agencies depend on revenue or other benefits received from the sale or provision of the data, the more we may expect to see agencies act in the ways described above.
d. The more secondary use of information occurs, the more that categorization systems and standards are likely to be used to influence the content, structure, and usefulness of the data to the secondary user.
e. Shared norms and expectations are also likely to shape whether and how much secondary users and agencies are likely to act in the ways described above, as follows:
    1. The more that agencies share a norm against provision or sale of data (in their own interests or those of other stakeholders in the data), the less likely we are to see the relationships described above.
    2. The more the benefit to the agency of selling or providing data, whether by secondary use being seen as furthering work related to its mission, or building support or legitimacy for the agency, the more we are likely to see these relationships.

3. Successful changes to policies on access to governmentally-held information are likely to be framed in a way that echoes previously established compromises and norms. This is because access policies are likely to reflect negotiation between public values that favor restrictions on access to data versus values that favor widespread public access to data. Attempted changes in these policies that could be construed to upset the compromise may attract attention and mobilize interests against the change. For this reason, secondary users and agencies are likely to try to make any changes appear to be less consequential than they in fact are.
4. The more the particular information and the financial arrangements regarding it are "submerged," with the provision of the data and its re-uses not easily visible to the public, the more likely secondary users are able to influence the content, structure, and availability of the data without interference by the press of public.
5. In any jurisdiction, agencies and secondary users are more likely to act in the ways described above to the extent that arrangements follow models that are widespread in other jurisdictions.

4) The content, structure, and availability of government data, as well as the relationship between government and secondary users of its data, are likely to be influenced by the interests of private sector intermediaries who assist in the marketing, provision, and / or selling of government data, in markets where they have come to exist.

a. Where intermediaries exist in the government/secondary user relationship for a particular type of data, we are likely to see "lock-in" and ongoing relationships reinforced by policy learning and sunk costs, as well as the benefits they provide by lengthening the "causal chain" and helping to reduce visibility and increase legitimacy of government benefiting from the provision of the data. This lock-in may also strengthen the ability to influence government policy in this area in ways that favor the intermediary due to government's dependence on these benefits as well as, potentially, the revenue it helps collect.

b. The intermediaries, in their capacity as marketers of data, are likely to influence government to package and/or shape the content and its methods of delivery to facilitate acquisition by secondary users, and to push for or develop new sources of data or data-related services that are more lucrative for both the intermediary and the government.

My expectations are summarized in simplified form in the following table:

Table 1 - Expectations

Government benefits or believes they can?	Industry produces revenue with data or believes they can?	Likely access policy / decision impact?	Likely data impact (content, format)?
No	No (low)*	No	No
No	Yes	Yes	No
Yes	No	Yes	Not possible
Yes	No	Yes	If possible

* If neither party had any benefit, there would be no relationship, so this case is one where there is some benefit to industry, but not critical to success of their business.

Research Design, Methods, Data and Data Analysis

My research design is a comparative case study of several different areas of private reuse of governmental data. My methods, to be discussed in more detail below, consisted of analysis of 250 documents and 65 in-depth interviews. The cases were selected to accentuate variation in *a) whether or not the government receives or believes they can receive benefits from reuse of their data; and b) whether or not industry produces or believes they can produce income from the data*. In general, I am also trying to understand within this framework what the barriers to or incentives for policy feedback are. I examined several policy areas involving specific types of data chosen based on the expectations above and involving different levels of government.

My rationale for looking at policy areas at both the state and federal level is that states have developed varying policies and practices on access to governmentally-held data. While some state-held data are made generally available to the public in easily-accessible formats, other data are made available only in specialized forms tailored to particular organizations or uses, or access to them is limited to certain audiences. In the case of the federal model, they are leading the push toward making data available for secondary use in machine-readable form with no restrictions on content or use and with no fee beyond the cost incurred to provide it (see OMB Policy A-25) where the agency chooses to charge it. For the most part, this means that when federal data are put online for access by secondary users, it is essentially free to them. And, while some state-held data are made available in a way that mirrors that model, other data are available only for fees that amount to profit for the agencies providing it. While the breadth of these variations is great, and there is no central catalog of initiatives or their terms from which to select, the following describes each case and my rationale for choosing it.

Research Design

State and Local Restaurant Inspection Records. *This is a case where a private company tried to get government to adopt a data standard that served the company's interests in generating revenue from the data. However, the government agencies that controlled the generation of the data gained no benefit from this private reuse and so did not facilitate the relationship.* The food safety oversight function under which Restaurant Inspections fall varies nationally as to whether it is conducted at the state or local level, or some combination – say, with large cities in a jurisdiction performing their own. The data here would not seem subject to influence, in that these are results of inspections performed according to a set of standards. Yet, these standards and even the way the results are presented to the public vary widely across the United States, and it is these results that are the subject of interest. The data is open, and other than some basic terms and conditions governments may enforce for their use, there appears to be limited policy guidance on access or distribution. And, there are vendors who produce revenue by crawling public websites to repackage this information. I have chosen this case specifically to focus on the history of an initiative by a company to encourage government to adopt a standardized approach to reporting that favored the company's interests in displaying this as part of a nationally available product. However, while they were interested and approached government about it, it was complementary but not core to their business model.

State Genealogical Records. *This is a case where private industry had a strong financial interest in the reuse of governmental data; although the government agency gained no direct revenue benefit from the reuse, it traded exclusive access rights to data to the private sector in return for services that fulfill its mission.* These records, held at state archives, are of interest to genealogists researching family history. Because they are in paper form, representing the output of business processes designed long ago, the data they contain is not subject to influence. In this case, the format of the record can indeed be changed – the subject of the case are agreements to digitize them, a process which changes them from paper to electronic format, and may also result in the addition of metadata. And, while access is not restricted and the paper copies are available to the public, as part of these same agreements, states grant a vendor exclusive rights to the resulting image for a period of time, producing revenue for the vendor. So, there is definitely flexibility in policy interpretation here, financial benefits to industry, and benefits in the form of digitization to the state, fulfilling the archives' mission of preserving the underlying record.

State Driver History Records (DHR). *This is a case where both industry and the state produce revenue from the private reuse of governmental data. While the nature of the data make it difficult to manipulate it to increase these revenue potentials, because government controls the supply, it has the potential to produce more revenue over time through raising the fee for the data, or implementing new products created from it.* As noted earlier in this chapter, driver history records (DHR) produce a large volume of revenue to industry as well as to states, although that can vary significantly based on the size and pricing in that state. It is likely one of, if not the largest, areas of data sale and revenue in the states nationally. Access to this data, however, is limited by federal and state law to particular uses, essentially creating a franchise for volume users to redistribute to the insurance industry. It also involves a large intermediary who, on contract in many states, brokers these transactions. So, there is significant revenue available to all three parties. The data would be very difficult to impact, given that it primarily consists of violations of the law reported by the courts, while there is flexibility in the format of the records. There are, as stated, policy constraints on use, but there are other decisions and policies that surround this subject, like pricing and products created from the data that allow for influence. See the Limitations section below for challenges encountered in research in this high-stakes, low visibility area.

Federal NOAA Data-Sharing. *Several cases examine different initiatives or programs within the National Oceanographic and Atmospheric Administration (NOAA) that reveal dynamics arising from variations in the degree of private-industry interest in income from governmentally-held data in relation to an agency that is forbidden by statute from generating revenue from private use of its data and is encouraged by statute and its mission to facilitate this* *private use.* NOAA is an organizational unit of the U.S. Department of Commerce. The NOAA cases primarily concern meteorological data provided by one of its subunits, the National Weather Service, and other bureaus within the organization. The cases are also interrelated, as I will explain below, and two involve new and novel programs. Studying these allowed for "real-time" insight into their evolution and decision-making, but, by definition, the full scope of their impacts have not yet played out. As NOAA weather information is one of the largest, and oldest, cases of open government data being used by industry and is instrumental to the production of billions of dollars in private sector revenue, my hope is that these cases provide evidence that is helpful in theory-building that can be used to better understand how or if influence may emerge in less mature or smaller scale open data markets.

NOAA Private Sector Environmental Data Sharing Policy. *This is a case where private industry has a strong financial interest in reuse of governmental data but the agency is forbidden to generate revenue from this reuse, yet has a mission to facilitate it. The case reveals how the private sector comes to influence NOAA data sharing policy to serve their interests.* Unique among federal agencies, NOAA has established a more detailed policy beyond federal open data and information management policies that deals directly with private sector use of their data. This case study examines the establishment of a policy framework by NOAA that results in the 2006 *Policy on Partnerships in the Provision of Environmental Information*, including private sector influence in that process, and subsequent efforts to shape access to NOAA data by some of the same companies involved later in the NOAA Big Data Project. This case, then, is a deep-dive into how industry attempts to and succeeds in influencing policy to its benefit in one of the nation's largest and longest standing commercial uses of government data.

NOAA National Center for Environmental Information Data Promotion Program. *This case examines a government program's attempt to increase awareness of and dependence on environmental data, to the point of potentially incorporating private sector requirements into data products, as well as using their knowledge of this data to try to jumpstart a new industry to deliver climate adaptation services in the service of their mission.* The NCEI, another subdivision of NOAA, and the self-titled ""Nation's Scorekeeper" regarding the trends and anomalies of weather and climate" (About Us, 2018) is the world's largest archive of environmental data. In recent years, a program has begun to promote the value of the data in the archive, specifically by studying and publicly highlighting dependence on it across multiple industries. This effort, conducted with an intermediary, goes beyond just marketing with the twin goals of incorporating more of the requirements of external users, including business, in its "products" and attempting to incentivize the development of a market for climate adaptation services in the United States to further its mission of increasing America's resilience in the face of climate change.

NOAA Big Data Project. *This is an unusual case in which an agency responded to a government-wide policy initiative to enhance private use of open environmental data for economic development and scientific innovation by initiating a "big data partnership" with industry. The partnership was based on a model the agency devised that they believed could meet this goal while also significantly reducing their own costs to deliver this data. But, NOAA's desire to structure the project with no cost to government while maintaining their own policy constraints requiring it to be free to the public and prohibiting "privileged access" that might favor one company over another resulted in them choosing private partners whose business models and incentives continue to make it difficult for either party to realize the intended benefits* *or mutual dependence. This case, then, follows the design and implementation of an ongoing multi-year project initiated by government to develop and benefit from increased private sector dependence on data, providing insights into both government and private sector incentives and barriers to mutual dependence*. In 2015, NOAA began to solicit industry for proposals to develop a relationship where NOAA data would be hosted and delivered to the public and business users free of charge by a private sector company. In turn, the company would subsidize that cost (a cost that was part of NOAA's mission, but that they would avoid going forward) by devising yet-to-be-defined value-added services the company could charge for. The design and implementation of this project provides a unique view of the exercise of interests and influence by both government and industry as they come together to negotiate a relationship of mutual dependence on government data. The data is open and the policy environment, while restricted by overarching federal policies, is somewhat mediated in its flexibility by the NOAA data partnership policy discussed earlier and the "no-cost" nature of the proposed relationship. While NOAA believes there is significant revenue for industry in the agreement, it was unproven at the start of the project. This case is related to all three of the other NOAA cases, as the eventual implementation involves NEXRAD data, some of the initiative is connected to efforts around the partnership policy, and the data and some of the participants were involved with both NCEI and their intermediary.

NOAA Next Generation Radar (NEXRAD) Data. *This is a case where private industry has a strong financial interest in the reuse of governmental data. Yet, while the federal agency is forbidden by statute from reaping revenue from the sale of its data, it is required by statute to make it publicly available and private reuse can help further its mission. While this data may be subject to construction, companies' interests are not in improving data that is freely available to* *their competitors, but in improving access to the data from which they create value-added, competing products. This case reveals the role of changes in technology in creating the playing field for influence as government and vendor relationships in data provision evolve over time, and how both parties make use of those changes to serve their interests.* Data from weather radars have a long history of dependence by industry, dating to the late 1940's, with users that stretch into many industries from agriculture to trucking and insurance. By looking at the evolution of its distribution through the lens of its "lifecycle" (from creation through distribution) over time, one can see the interplay of technology with agency mission and capabilities, along with the interplay of government and industry in its distribution. This data is open and while format and policy are well-established, there are still opportunities for changes as new technology and software algorithms are introduced. These data are also directly related to the NOAA Big Data Project below as the evolution of dependence with technology analyzed in this case culminates in it becoming the first type of data successfully deployed in that project's new distribution scheme. As noted in my research design, however, the mission of the weather radar function is public safety and direct access to its data by the private sector is important, but viewed as "nice to have" given that it is government's role to issue warnings. The following table summarizes these cases in terms of the research design outlined above:

Table 2 - Research Expectations by Case

Case	Type	Government benefits or believes they can?	Industry produces revenue with data?	Likely access policy / decision impact?	Likely data impact (content, format)?
Restaurant Inspection	Data	No	No (low)*	No	No
Genealogy Records	Data	Yes	Yes	Yes	Not possible
Driver History Records	Data				
NOAA Environmental Data Provision Policy	Policy	No	Yes	Yes	No
NOAA Data Promotion	Program	Yes	No	Yes	If possible
NOAA Big Data Project	Project	Yes	No	Yes	If possible
NOAA NEXRAD Radar	Data	No	Yes	Yes	No

* If neither party had any benefit, there would be no relationship, so this case is one where there is some benefit to industry, but not critical to success of their business model.

Methods and Data

Research Method. The methodology I have employed in this study is qualitative, involving semi-structured interviews, email correspondence, and document review and analysis. A qualitative approach is most suited to this research topic because much of the work is exploratory, seeking to understand motivations, incentives and mechanisms for exercising influence, decision making, and the understanding held by participants of markets and opportunities for revenue, along with barriers. While some quantitative data exists on this topic, say, on projected costs, or volume of use of particular records or data, some of which is included in this research for descriptive purposes, many of the measures associated with the private use of data are proprietary, or, in the case of the Big Data Project, protected by a non-disclosure agreement, making it challenging to develop an effective quantitative framework for use in analyzing this subject.

Participants. The interview participants interviewed in this study were, for the most part, directly involved either at some point, or currently, in the cases under examination. In one occasion, quotations were taken from a recorded presentation that was presented publicly at a national conference. And, there is an ethnographic vignette included in the study of genealogical records that includes comments from participants in open meetings who had been informed under human subjects of my research purpose and their rights.

Participant Selection and Recruitment. Participants in the various cases, described generally below, were recruited as they were identified as key parties involved in a particular business function or project via websites, online documentation, or through contact with knowledgeable parties in government or industry who referred to experience or involvement of other individuals. Identification also included use of advanced features of LinkedIn professional networking software to identify and solicit participants, including retired employees and representatives from private sector companies.

Interview Instruments and Protocols. All individuals in the study were verbally read and agreed to an Information Statement outlining the purpose of the study and their rights as a research subject that was approved by the Human Research Protection Program (HRPP) at the University of Kansas, Lawrence Campus. Questions for each semi-structured interview were customized based on the role of the participant in the function or project and other research from documents or policies they may have participated in crafting or meetings they attended or presented at, but fell within the general purpose of soliciting answers to the research questions approved by the HRPP for this study. All interviews were taped and transcribed verbatim in their entirety, with printed transcripts produced for each. Almost all interviews were conducted by phone, usually of just one person, with two interviews using desktop videoconferencing software, one in which both a company representative and the company's attorney participated as an observer / advisor. Occasionally, supplementary and follow up exchanges to the main interviews occurred via email or LinkedIn messaging. Limitations on time and funding prevented travel to participant sites which were located around the nation.

Level of Participation / Number of Interviews and Resulting Documentation. Across the cases included in my dissertation, I conducted semi-structured interviews with over 65 individuals with an average interview length of approximately one hour, producing over 1,200 pages of documentation in verbatim transcriptions. This total does not include comments and interviews associated with the ethnographic portion of the case study on genealogical records, which was made up of several individual interviews along with field notes and recordings made at several meetings and a professional conference. In addition, across these seven cases, I reviewed and analyzed over 250 documents included in the bibliography accompanying this study along with reviewing at least another 50-100 in the course of my research, including state and federal statutes, policies, white papers, requests for information (RFI) and requests for proposal (RFP), formal government reports – including audits, contracts / agreements, press releases by industry and government, numerous slide decks from public presentations at conferences and before professional groups. I also visited and reviewed numerous websites and made extensive use of the Internet Archive's "Wayback Machine" to review past copies of webpages and materials that are no longer available on government and / or private sector websites and listened to audio recording of legislative testimony.

Methods of Analysis. I analyzed interviews and documents using my research questions, looking for patterns and evidence that supported or contradicted the expectations outlined above. I also used these sources to identify the interests of government and industry in each case, looking for barriers and incentives to feedback which I then compared across cases to look for commonalities and differences in these factors and their impacts. Because all of the cases involve a "dialog" between industry and government in some way, by talking with people who participated in different ways in the situations described in these cases, sometimes at different stages, I could “match them up” by varying some of my interview questions. This provided data about both the evolution of the situations and relationships described, and also how the participants viewed each other’s motivations and actions, say, around a policy or opportunity, which provided support for my analysis of influence.

Possible Research Bias. As an employee in state government, I have some experience in the mechanism used for the sale of driver history records in Kansas, as I have served as a representative of a board that oversees the work of an intermediary during my career. In turn, at least part of the time I was studying the driver history record area, I worked for that board, but not in a role that required me to supervise this work. In addition, I have experience with technology in a government environment in my previous work roles as a project manager / director. Because of this, it is possible that my questions or perspective on this topic may have become biased by this experience. The continuum of this bias would presumably range from omitting questions that might poorly reflect on government behavior or mechanisms of influence to doing the same with the vendor or intermediary community. Because the evidence I present does not cast any of these parties in a light that fails to identify their influence, and I was challenged to find contacts in the vendor community to interview, I feel any bias was successfully minimized. In the case of the case study on the archives, I served for a number of years on an oversight committee in this area. Like with driver history records, this gives me insight into the mechanics of similar operations. In the case of genealogical records, however, my thesis involves ethnographic work, interviews, and document review that suggests the possibility for influence to occur that negatively impacts their mission, again, countervailing the idea that I would hold a professional bias in favor of minimizing their role in mutual dependence.

Limitations

State Driver History Records

In the course of my research, I discovered that besides wide variation in the state organizations responsible for administering theses sales, the locus of control for policy decisions was relatively opaque, with turnover making the administrative history of decisions and their influences difficult to trace on the state side. In turn, private sector representatives of data purchasers did not make themselves available for interviews, so my work in this area consisted of interviews with an opportunistic sample of managers from branches of state intermediaries in data sales, and a few state representatives with knowledge of the history in a state. I also attended and listened to hearings of the legislation referenced in the case and reviewed related documentation.

Contribution to the Literature

To my knowledge, while, as previously referenced, there have been numerous studies of open data, there have been none on the relationships, behavior, and impact of driver history record sales at the state level, nor of NEXRAD weather radar, the NCEI Data Promotion efforts, nor the NOAA Big Data Project. So, much of the domain areas are newly researched, especially from a public administration perspective. Policy Feedback Theory has, from my review of the literature, not yet been used in a single research study to assess the phenomenon of government information as a resource / benefit. So, it is my hope that the findings from this study may lay the groundwork for more research into the political and bureaucratic impacts of commercial use of government information. And, in my recommendations I offer suggestions for how this area can be better mapped to provide visibility into cases for that research.

Chapter Outline

Chapter 2 includes case studies of sale of State Driver History Records (DHR), State Genealogical records held in state archives, and Restaurant Inspection records held variously by both state and local government. Chapter 3 addresses the case of NOAA Environmental Data Policy and the NCEI Data Promotion efforts. Chapter 4 provides an analysis of the conceptualization and contracting phase of the NOAA Big Data Project, with an exploration of its execution continuing in Chapter 5. Chapter 6 analyzes the evolution of the distribution of NEXRAD Weather Radar Data. The study concludes with Chapter 7 which discusses this study's findings in light of the research questions and policy feedback theory as it applies to government information as an economic resource. It proposes direction for future research and potential implications for policy and the profession.

Chapter 2: What's in it for me? Three Cases of Industry Dependence on State and Local Government Data and their Impact on Data, Policy, and Practice

Introduction

We begin the study of influence in the secondary use of information with three case studies that illustrate the progressive effects of government dependence on private use of its data and their impact – realized and potential – on government data, policy, and practice. In the first case, we see a company become interested in government data about Restaurant Inspections. This is open data, essentially free for the taking, but held variously at the state and local level and made available across hundreds of their websites across the nation. To make the data cheapest and easiest for the company to use as an added attraction to their product - a national online service directory that includes restaurants and ratings - they want government to put the data in a standard format and they launch a national initiative to accomplish this. It turns out, however, that government sees little benefit - the inspection data is already online. And, they are pulled in the opposite direction by the restaurants they regulate (not big fans of increased transparency), and by staff who see infrequent "point in time" inspection results as potentially misleading to the public, and perhaps revelatory of inconsistency in their own practices. In the end, the company becomes frustrated and hires another company to pull and process the data after the fact. It is not critical to their business model, so good enough. From this case, we see the challenges to building dependence when industry incentives are not strongly connected to revenue, but also where government perceives little benefit. Along the way, we gain our first insight into internal factors that work against responsiveness to industry demands, the ways industry responds, and an ecosystem of other users and business models that can impact the eventual outcome.

The second case concerns data held in government archives that are of interest to a company whose business model depends on wide public interest in genealogy to sell access to these records online. In this situation, government archives have a problem that industry can help them address. The records are old (non-current, thus their residence in an archive) and exist almost exclusively on paper, and are thus subject to degradation in quality over time. They also take up significant space to store (with associated costs), while limiting public access by requiring physical trips to the archive – both gaps in major elements of the archives' mission and ones for which they have little funding to address. The company steps in to fill this void, offering services to convert these records to digital format for no charge. The catch? The archives must agree to grant the company exclusive rights to deliver this data to the public for a certain time period. The archives willingly agree, addressing the issue of the fact they are open records by gaining an exception for their citizens to have access free of charge from the start via the company's website. But, several aspects of the situation give pause for thought about the impact of this new mutual dependence. First, while there are many records of interest to the public awaiting digitization, the market moves these projects to the front of the line. As funding challenges continue, could this shape the nature of the history that, in practice, is available to the public based on market values? With little funding, when the period of exclusive access expires, the archive will not likely to deploy the technical infrastructure itself to open up access to everyone. At the end of this section, via an ethnographic vignette of decision-making about whether to retain or dispose of a type of record that might be of interest to industry, we are given the chance to assess the factors that might lead to another, even more significant impact – that the value of the market for use of records might seep into decisions about which records are even retained by government, a decision critical to the functioning of transparency and democracy.

The final case in this section represents an even more developed exchange of benefits in the form of money – and lots of it. Here we visit the case identified early in the introduction to this study, where industry – multiple companies – are able to buy and sell individual driving records held by states, producing significant revenue for themselves and to state government. In this case, the government benefits from this revenue, to the tune of millions of dollars, and these funds have sometimes come to support government functions unrelated to the production or internal use of the data. While federal and state policy limits the use (and thus users) of these records, and the data is mostly static, reflecting violations of the law, we witness a variety of ways this mutual dependence impacts government as they show a willingness to protect threats to this revenue and its use from market substitutes, competition among agencies, and even legislation. We also gain some visibility into the role of a private intermediary that has moved into this "niche" (Pierson, 1993), contracting with government to provide support in these relationships by handling the data distribution and relations with vendors, further complicating the influences involved.

By the completion of this chapter, the reader should gain a greater understanding of what well might be an area of government previously unknown to them, and the arc of how increasing government benefits from the commercial use of its data may come to influence the data, policy, and practices of government in ways that challenge current paradigms of transparency and good government.

Restaurant Inspection Data

Elizabeth was in the passenger seat as she and her husband neared the last moderate size town for miles on their interstate journey. She peered down at her smartphone screen in the dark, flipping through a list of restaurants. As they looked for closing times – it was already late – she noticed a small icon in one of the entries that listed a score for the "health" of the restaurant they were discussing and then realized they all must show that. "Yeah, Dave's Café looks great" she said sarcastically, “It got a 71 out of 100.” She continued to squint in the darkness. The most recent inspection showed no violations, but there was no more detail. She looked briefly for a way to filter and sort the restaurants by this score, but while “Good for Kids” was one of the selections, this score was not. After some more conversation, they picked one of the other restaurants that was closer to the highway anyway, with an equal number of good reviews, but one they also now knew had a health rating of 88.

This anecdote is partly-based on real-life experience, but not far from what the head of a state agency’s food inspection division has come to expect with regard to public access to information about their periodic restaurant inspections:

> I don't think the average person's going to... that is driving through (our state)… wants to find out who in this state is in charge of doing inspections, then go to that agency's website and find the inspections to actually look at a restaurant. They are going to use Yelp or anything else, social media, to actually determine the actual best restaurants in a city. I'm not naïve enough to think people are using our website to actually make dinner plans. (PRI-01)

Food inspections, especially those focused on food handling practices at restaurants, are a critical behind the scenes function of government in support of public health and the prevention of outbreaks of food-borne illnesses. In one Midwestern state, a workforce of 50 home-based inspectors performs roughly 23,000 inspections each year across a population of 18,000 licensed food establishments that include grocery stores, restaurants, convenience stores, schools, mobile food units, along with food processing plants, food storage facilities, and lodging establishments (PRI-01). The problems these inspections are designed to prevent can cause significant damage to both the lives of individuals and to commerce. In fall 2015, for example, 55 cases of foodborne illness across 11 states were reported from eating at Chipotle restaurants, causing a temporary nationwide shutdown in November 2105 (FDA & CDC, 2016). Out of 839 outbreaks (defined as two or more individuals getting sick) of foodborne disease reported to the Federal

Centers for Disease Control and Prevention in 2016, for example, there were 14,259 illnesses and 17 deaths, with 363 of the outbreaks coming from a restaurant with sit down dining (CDC, 2106b).

The following story provides an abbreviated view into an ecosystem of distribution, policy, and technology involving government information – Restaurant Inspection data. It examines an attempt to implement a standard for distribution of data from these inspections that involves multiple companies who in some way depend on the data for commercial gain through its transformation into an indicator that is made broadly available to the public by a private sector company.

Data, Distribution, and Organizational Structure: Barriers to Commercial Reuse

Several aspects of the situation in which restaurant inspection data comes into being are important for understanding both the shape its commercial use has taken and the resulting strength / weakness of manifestations of both dependence and influence. **The first is that state food safety programs are not all administered the same way.** While in 38 states, a single agency regulates retail food stores and restaurants, the others are split across multiple agencies (FDA, 2018). When Washington, DC is included, this number totals 65 agencies in all. **Second, the inspections are not all conducted using the same rules.** The FDA creates and promotes a model food safety code that "assists food control jurisdictions at all levels of government by providing them with a scientifically sound technical and legal basis for regulating the retail and food service segment of the industry…" (FDA, 2019). This code has moved over time from being updated every two years to a four-year cycle (FDA, 2019). However, to the degree that states are uniform in adoption, they are far from uniform in the <u>version</u> adopted. In 2018, states were on seven different versions of the FDA food code, beginning with one state still on the

1995 version (the most current version is 2017). As the executive at a state food safety program cited above stated in 2019:

> [T]here are certain states, or certain jurisdictions that it is a big undertaking to actually adopt the newest version. The actual model food code comes up once every four years. We are in the process currently of adopting the 2017 model FDA food code. And before that we had adopted the 2009 version. And we had that ever since 2012. So, it took three years for us to adopt the actual 2009 version. (PRI-01)

So, across the U.S., as of 2018 we have 65 agencies operating on seven different versions of food safety codes, with California adopting their own (FDA, 2018). **However, from the standpoint of comparability and consistency in the data produced by these inspections, a third, even more critical aspect comes to the fore: The states (and some cities) use different rating scales to present the results of the inspections.** In a 2014 survey of by the National Association of County and City Health Officials (NACCHO) of retail food inspection and rating systems, they estimated that about 2,100 of the 2,800 local health departments in the U.S. are "educating, inspecting, or licensing food retail establishments," representing about 70% of the 3,000 entities regulating food safety nationally (NACCHO, 2014). In that survey, about 75% of those respondents having a Food Inspection and Grading system reported using a numerical score, 16.5% used letter grades, 10% a color or graphic, and 11% some other approach (with 16% using more than one) (NACCHO, 2014). Among those using a system at the time, 37% made the rating available on the Internet (NACCHO, 2014). While the response size did not produce a 95% confidence level in the results, they illustrate that significant variation exists.

As one might imagine, this variation introduces challenges to restaurants and the public as well in understanding the application of rules and the meaning – and comparison – of results. While county borders matter little to the hungry family – and to the traveling public - one can quickly see cases where this might become a problem not just for comparing the practices of restaurants, but for the restaurants themselves in standardization of practices and training efficiencies. As the state executive over restaurant inspection recounted: “I was talking to one of the senior members of McDonalds, they have stores in the same town that are a mile apart. But, they are in different counties and they have different food codes and different rules to follow in that mile apart” (PRI-01).

Influence and the Need for Government Standardization in the Private Sector Distribution of Restaurant Inspection Data. From the perspective of a commercial company with a national scope wanting to make this data available to consumers, these variations pose several obstacles. In the weather industry, for example, to make the local temperature outside available through a smartphone app, one need only provide a single, standardized figure – albeit for a nearby location (the National Weather Service doesn't track temperature for every square foot of the country). But, in this case, the inconsistency in the presentation of the ratings for food inspections poses a problem for streamlining both the presentation of data and its acquisition. For an interface on a mobile phone, there is little room on the screen to provide context between numbers on different scales (is it 4 out of 5, or out of 100? How does that compare to a “B”?). To continue the previous analogy, it is almost as if some places recorded “92 degrees” – others only “hot” or “pretty hot” – some the “heat index” and others just the strength of the impact (severity), like “could cause heat exhaustion.” In that vein, for example, in at least one state, no score or grade is calculated, with only the number of violations at each level of severity reported (PRI-01). Next, to the degree that multiple food codes are being used, likely across different frequencies of inspection, even when a number or grade or symbol is presented, its meaning may not be obvious for a particular health concern. When added to the administrative aspects of acquiring the data – multiple points of contact in some states, including cities, and the fact that, as of several years ago less than half the data was online, likely sitting in systems using different software, and managed with varying levels of automation – implies a significant effort to obtain it in electronic form, on the timely and recurring basis needed by industry. Enter the Local Inspector Value-Entry Specification (LIVES) (Yelp, 2019).

From Supply to Demand: Commercial Efforts to Implement a National Specification for Restaurant Inspection. The LIVES is a standard intended for government use in sharing restaurant inspection data with Yelp, a large company that provides directory services, including forums for reviews, both online and as a smartphone app for restaurants and other businesses to a market of millions of customers across the United States. The standard was developed in 2012 and "pitched to government and branded to the outside world" (PRI-02) as a method for standardizing the presentation of restaurant inspection data and delivering it to a larger audience, "saving lives" through raising public awareness of risk of contracting foodborne illness at restaurants and incentivizing them to improve their sanitary practices (PRI-02). The first implementation was in a large city, promoted in a press release by the mayor, expected to be followed by some other cities nationally. According to a company representative who confirmed that he was only speaking for himself, not the company during the interview quoted extensively below, they were headquartered in a high-tech city, so there were a "bunch of technology geeks that actually work inside City Hall," making it easy to partner with them to launch the service (PRI-02).

Government Benefits versus Private Sector Interests in Cooking Raw Data

On the outside, the challenges posed by the variation in government's presentation of restaurant inspection results, the food code that is applied, and the diverse and distributed organizational structures that support this activity have been made clear. And, yet, somehow, the data covering a large swath of the United States is now being used to present a standardized restaurant health rating. The implication may be that getting state and local government to standardize their reporting in service of a larger commercial - or even public - purpose has largely succeeded. As illustrated in the anecdote that opens this section, "health ratings" covering a full set of restaurants in at least 20 states, with partial coverage in at least 17 more (Yelp, 2019b), are now delivered through this and at least one other commercial smartphone app.

However, the situation – and what constitutes success - is much more complicated. This new request is seen through the lens of past efforts and investments that already required the data to be available locally. Employees and managers push back on greater access, concerned that periodic inspection data is a misleading indicator for the public, and further access would only amplify the confusion. It also seemed it would introduce greater scrutiny to their practices, along with accountability. In turn, restaurants, another industry who has a stake in the data and much closer relationship to the inspectors, were not supportive for the same reason. We later see an instance where industry is unhappy with concessions made to the restaurants once the effort is underway and are willing to use the press to bring pressure on both parties to fall in line with their purposes. The specification is eventually adopted here and there as more barriers are revealed, including the private sector's lack of understanding of the government barriers discussed above and the purely altruistic motivations they first impute to these decision makers, only to later uncover the hard reality.

We begin the story with an account from an official working in public health at the city at the time the standard was proposed, then cycle through the perspectives of other parties involved in the project to create a portrait in miniature of the interests, incentives, and challenges that characterized these efforts to commercialize restaurant inspection data. The level of detail parallels the in-depth conversations recounted later in the NOAA Big Data Project analysis, demonstrating the complexity latent in what could easily seem like just a technology matter. This first example, too, situates this new effort to make government data available in the context a series of previous efforts focused on this and other datasets, with the politics of restaurant inspection shaping government's initial engagement of the standard, and the incentives and economic dependencies of the commercial intermediary on the same restaurants – and government - combined. A city public health official described her first encounter:

> Basically, [the head] of our Office of Innovation asked me to participate in ongoing work where [an executive from the company] and a gentleman from (another large city), I guess, Office of Innovation who had been working on a single data standard for restaurant inspection. So, there was an initial draft of a data standard and I was presented with the opportunity to go in along with (the other city) to be kind of a pilot for the standard as well as the implementation via the company.
>
> From my own perspective as responsible for environmental health and maybe just like a little bit of an unconventional bureaucrat - or not a bureaucrat - in a bureaucratic position, I thought that this was advantageous because, pretty simply put, more public transparency of restaurant inspection results...would be a force that led to restaurants being more proactive in their efforts to meet standards. And internally for the health department, more transparency, more, I guess, efficiency and accountability internally.
>
> So that was pretty set. What I understood from the company was that their pain point was that every health department had different data structures, data systems, and data parameters. And, in order for it to make sense, they had to have kind of a universally consumable data format. (PRI-03)

This request was, however, delivered in the context of previous efforts to improve transparency of restaurant inspections. The official related part of the history, shedding light on some of the reluctance he encountered to delivering the data through a commercial firm, illustrating how history and relationships in both delivery of data and the commercial ecosystem that surrounds it can be factors in how distribution, dependence, and influence are shaped:

> So, that's when I jumped in. There was really no one in the restaurant part of the health department that was motivated to participate. And, when I talked to my subordinates, they were, frankly, reluctant...Again, this was not new to me because of earlier efforts at local transparency driven by external forces that had already occurred...And I think this is

> relevant. The legislature made restaurant inspection results publicly available by law in 2004 or 2005. The first consumers were the local newspaper. When they asked for the data in 2004/2005, I asked my staff to produce it. And they were basically unwilling to produce it, and the kinds of responses that I got from them were that people won't understand this. The two major themes were that the public won't understand it and the restaurants will be upset. So, I asked my IT team who was a little less beholden to the culture of environmental health restaurant inspection to extract the data and look at it.
>
> In our city, a compromise from an earlier initiative was that we would not have letter grades, we would make the restaurants put up their restaurant inspection report, the paper report. And there would be a score, zero to a hundred, and that would be not posted, but put on a card available for review. So, we implemented that law, that version of the law, and the local newspaper asked and then there were other "askers" so we regularly uploaded the data once a week to a file share. And when we were doing that, we were doing that right away. But, of course, it was just our data and our standards, not the LIVES standard.
>
> (But), when we looked at it, it basically showed that the inspections weren't being done at the frequency that was stated. So, it was basically, "we need more staff because we need to do three or four inspections." And here when you looked at it, there was a wide variety of performance between different staff members, some districts getting less than one a year, some doing two or three in a year. So that ended up leading me to kind of an internal accountability effort to figure out what that was about and trying to correct it. So, that was a little bit of the background.
>
> So, with the request to participate in the standard, I think the same kind of reluctance appeared. I think the restaurant inspectors saw the restaurants also as their constituents in part because the restaurants pay the bill, pay their salaries - indirectly [fees are charged for inspections] - I think that's a component. (PRI-03)

This idea of dependence on the regulated party, restaurants – who actually fund this work through fees - also came across in comments by the head of a state food inspection division referenced earlier: "One thing that the state takes seriously is educating and helping our customers, which are the business owners" (PRI-01). And, later:

> My strong feeling is at the state level, we are solely fee-funded. We do not get any state general funds. So, the fees we charge our customers have to pay full for the program. And working with stake holder groups...We have come to a determination based upon size and risk of the actual number of inspections we do. (PRI-01)

This orientation toward cooperation in achieving objectives is not unique to these two examples. The national Association of Food and Drug Officials (AFDO), with a core membership of government officials, has launched “Partners with a Common Purpose,” with this purpose summarized in the initial announcement:

> Within this collaborative initiative, AFDO will begin with food protection professionals and regulatory officials who will engage with each other in a “safe-harbor” environment to examine their ability to impact food safety control through discussion, self-examination, and forums. All participants will be equal partners and their discussions might involve characteristics of successful programs, barriers in implementing intervention strategies, lessons learned, sharing of best practices, and future opportunities. (AFDO, 2017)

It seems clear that while wide distribution of restaurant inspection scores might raise awareness and result in consumer-behavior driven compliance, a “collaborative” approach with industry might be seen as more preferable by health inspection departments.

In the end, with the city it fell to the executive in the public health department to implement the standard, based both on her experience with the data and the resistance she faced:

> Thus, (for the implementation of the standard) basically it was me doing the work, sitting and writing the standard, doing data extraction, etc. because there was little interest from the staff, either the line staff or the management staff in the restaurant inspection division to participate. [But] we had the data, I got standards, I got where they were going and I thought that the further visibility that the company could offer would be a cleansing for us. (PRI-03)

One takeaway, then, is that far from seeing this partnership as a way to extend and accomplish their mission by publicizing their work to a wider audience, there was internal resistance to doing so. Thus, a key factor in dependence is missing: The staff of the program creating the data did not see a benefit, even if this executive did.

The restaurants, of course, had their own perspective. In this case, the data was regulatory information to be provided to a national company that, among other services, provided the capability for consumers to post restaurant reviews. So, they had a little history with the company. The executive continued:

I did notify the restaurant association. I told them because I had a good relationship with

> their kind of public liaison, government liaison person. He was <u>not</u> "not supportive," but he felt that the board members, basically restaurant owners would not be supportive.
>
> There was also, I think, a particular fear or antipathy to Yelp because some restaurant owners had perceived them as being mercenary or unfair, and they thought that the data might be used to extort a restaurant in some way. Some restaurants perceived, rightly or wrongly, and I had no way to judge it, that - for a fee - basically they would increase your score. And if you didn't play, you'd be in the dog house. Or would they make the scores prominent or something like that. (PRI-03)

While the executive may have lacked evidence, this concern persists even to this day. One important aspect of this situation is that the smartphone application and website are free to use – the revenue (and thus a point of dependence for the company) is derived from advertising on the site that is sold to the companies, <u>in this case restaurants</u>, being reviewed. In 2014, the Court of Appeals for the Ninth Circuit found that the site could legally "lower or raise the rating of a business depending on whether it advertises with the company" (Associated Press, 2014). A academic article analyzing the review filter on the website around this time (Kamerer, 2014), cited almost 700 complaints filed against the company with the FTC, along with newspaper articles and even a TV episode, all emphasizing that "Many small business owners claim that Yelp uses the review filter to reward advertisers and to punish everyone else" (Kamerer, 2014). Yet, efforts to implement the standard moved forward.

Finally, in January 2013, the new specification went live on the Yelp website. The following excerpts tell the city's, and the company's, story:

> **WASHINGTON, Jan. 17, 2013 /PRNewswire/** -- Today Mayor Edwin M. Lee, Chairman of the US Conference of Mayors Technology and Innovation Task Force, and Yelp CEO and Co-founder Jeremy Stoppelman announced the initial integration of city-provided restaurant health score information on the site that connects people with great local businesses. San Francisco will lead the charge on this innovative effort to make valuable government data more easily accessible to the public; New York City restaurant grades will also be added as business attributes in the weeks ahead….
>
> …"This new partnership with Yelp to offer restaurant health inspection scores on its site is another significant step in the Open Data movement," said Mayor Lee. "By making

> often hard-to-find government information more widely available to innovative companies like Yelp, we can make government more transparent and improve public health outcomes for our residents through the power of Technology..." (Yelp, 2013)

A critical element of the press release underpinned the proposed incentive for participation by government health departments:

> According to a study in the Journal of Environmental Health (March 2005), Los Angeles County's decision to require restaurants to display hygiene grade cards on their entrances led to a 13 percent decrease in hospitalizations due to food borne illness. The study also demonstrated that the mandatory public display of these health grades improved the overall average score of restaurants in Los Angeles by incentivizing improved best practices across the local industry. As a leading website and app for dining decisions, Yelp's open data initiative LIVES stands to empower consumers and improve the quality of life within the cities that participate in the program. (Yelp, 2013)

Promoting Adoption of the Specification: Influence and Altruism Encounter Administration. It turned out that the other large city dropped out and so this city went forward alone. We begin the story of the rollout of the standard with the perspective of a participant from the company:

> So, we announced it and I think our plan was naive in that we expected every municipality - it's usually municipal governments that manage this stuff - to form a line and begin entering their data exactly as it was prescribed in the spec. And we'd begin ingesting and just sort them according to population. And then within a couple of years we would have everybody – most of the U.S. population covered.
>
> And of course, not only did that not happen, it became pretty obvious pretty quickly. I think we launched it at the U.S. Conference of Mayors. And anyway, it just became obvious that unless you force feed it to them, there's not going to be a way to get a bunch of municipalities on the same page about even really basic non-controversial things when it comes to data sharing. And I think good example is <the other large city>.
>
> I mean, we had all the ... And, in fact, there was a technologist in the <large city> who helped us design the spec, and gave us some ideas for how it might work for <them>. And then when it came time to actually bring their city online, the Environmental Health Department vetoed it and said, "We don't want anything to do with it," because they interpreted having even a numeric score that wasn't visible to consumers powering the backend was a problem. They just had their ABC thing, they didn't want anything that reflected a numeric attribute for their system because it was ABC's [a letter grade]. And so they didn't sign on to it.

> So, with the city we started with, it was easier for them because we really just called them up and said “Hey, tell us about how you do your stuff.” And that’s where we got the zero-to-one hundred scoring. And that makes sense, too, assuming that is also a numeric quantitative thing that’s not wildly – it’s not one-to-five star, which they probably do. (PRI-02).

In these statements, we see an unambiguous attempt by a company to influence government in the form of enlisting their participation in a standard, but one only used by the company, for public relations and, perhaps, a larger health purpose. The appeal, as we see in the press release above, was based on the ability to inform a larger audience and leverage the findings of an empirical study that found a significant correlation between raising public awareness of inspection outcomes and lowered instance of foodborne illness to improve a city’s health (Simon, 2005). However, we see also that there is some promotional value for the city.

In this city, attempts at influence did not end with their adoption of the standard, nor did the forces and incentives leading government and the restaurants to resist disappear. Witness the following anecdote:

> We started putting health alerts on the pages of businesses in the lowest 5% of hygiene inspections saying, "Hey, this restaurant has not been ... It was deemed in the lowest 5% or whatever." So, the city health inspectors actually created a pilot program that allowed - our policy was basically, leave the alert up for, I think it was six months because that's how long it took before you get a new grade - and they piloted a new program that allowed you to basically pay to get a re-inspection and regraded. Imagine flunking a pop quiz and being able to take it the next day!
>
> And, so, then we changed the pop-up alert. We had to change the code of the pop-up alert because the city was basically allowing its restaurants to trick our system because the pop-up would be disabled as soon as we got a new grade. And so we overrode that on our backend and then put a note that said, essentially “This city is basically creating a loophole that lets businesses pay a little bit of a fee and get these alerts taken off. Call them to let them know what you think.” (PRI-02)

He continued, recounting how informing the local news department in the city could also help bring pressure. Sure enough, there’s a story titled "Restaurants with Poor Health

Ratings Can Now Buy Opportunity for New Scores," with the subheading "A new pilot program allows poorly-rated restaurants to purchase an “inspection do-over” in hopes of earning a better health score" (Shaban et. al, 2017). An excerpt from the news report, quoting one of the supervisors of the city restaurant inspection program:

> She says websites like Yelp’s make it easy for consumers to find a restaurant’s bad health score, which is why, she said, restaurant owners pushed the city to offer a re-scoring option.

He continued:

> And so the adversarial nature is not only with the restaurant community, but I mean, with environmental health inspectors as well. Because it's like you're dealing with almost two different entities. There's the city and the mayor, and the mayor typically appoints the Chief Technology Officer. So, you have these pro-technology cheerleaders. And then you've got these career people who are not political at all, and it's their way or the highway. Environmental health inspectors that have been doing it for 20 years, and understand their quirky systems, and they're not going to change it. (PRI-02)

He went on to say that he thought the inspectors were, by the nature of their job, close with the restaurant industry, attending the same conferences and developing their practices together, adding “And I think at about that time is when we were like, "This is taking too long." The old vision of the LIVES standard was to save lives” (PRI-02).

The representative from the company then related his subsequent experience in attempting to enlist participation nationally, along with his view of the reason why units of government should want to participate. The resulting discussion highlights not only dimensions of public information ecosystems, motivations, and potential dependence, but ones that might apply to other types of regulatory-related data. After the initial roll-out:

> I went to all these major metros, hat in hand, for what seemed like a really good idea. Take your stuff that you ostensibly wanted to have in front of more consumers, so let's get it on. You just have to adhere to the specifications, that's pretty easily customizable.
>
> And they all sort of balked because basically everybody thinks their system's the best. And I think quietly, our vision was to get as many people online as possible so we can

begin doing city-by-city comparisons…Like to see if there were particular metros whose scores maybe skewed more positively, implying that there's more capture between the restaurant industry and the inspectors and so forth.

I mean, that's the big thing - I think that it really is that consumers don't have a seat at the table for these. With this kind of information, the grading, all of the data itself, even if you put the data online in an open data feed, much of the raw data, if you're reading it - looking at one of these inspection forms as somebody just trying to figure out where to eat - it's not consumer friendly at all.

And, so, it might mean something to the restaurateur, it might mean something to the environmental health inspector. But if the point of the regime is to minimize food poisoning among your population, and the point of open data is to increase transparency - those are two laudable goals, but there's this bridge of how do you put this into a place where people ... Actually, how do you put this into a kind of a package, a presentation that makes sense to Jane Consumer, and then also catch them at a time when they haven't just pulled up to the restaurant, they're actually about to from in the middle, making their dining decisions? (PRI-02)

Returning for a moment to our opening anecdote – where the "Jane Consumer" just referenced failed to notice the restaurant health ratings at first - and in line with the recurring altruism expressed by the company representative in wanting to improve health outcomes and save lives, another aspect of the situation comes out:

The other funny thing is I think that a lot of consumers don't actually notice the grade because it's such a diminutive feature. So, yeah, we do some stuff for the worst offenders, like the popups. And it is more conspicuous on a smart phone or on the app. But what's interesting is I think probably most consumers don't notice it. But that doesn't mean that it doesn't have the impact. Because businesses are obsessed with their pages (featured in the application). They're looking at them several times a day, at all minutiae. They're looking at every pixel, they're psychoanalyzing every reviewer, any of their ... They sort of live and die by it. So even having the number on there, if the businesses get a low score, then they're telling their kitchen staff, "We've got to do better. We've got to be cleaner." And I think it's working. (PRI-02)

In the end it was like, I thought if we build it they will come. And it was like, instead we were having to beg people to come on. And then lots of inbound interest in governments putting their data online. But it's from random places with tiny populations, or places where there just happens to be a really geeky CTO, but not necessarily a big population so the bang for the buck is low.

I can't remember the exact jurisdictions, but some approach us and then there's no technical chops internally. And so they don't know how to even build the data feed. And then in the end it was just like, we're just going to go partner with a scraper and just go

> get this data. And they can put it into the spec and they be can sort of, we're an intermediary, but they can be an intermediary between us and the governments.
>
> Anyway, so we got just frustrated. We sort of waved them off in the past because we always preferred the idea of having a pure mainline direct feed from municipalities. This is something we don't make money from. I don't think it's necessarily a competitive advantage. It's something we have to invest engineering resources in. And we really see a little bit of a credit on the PR side. But it's just something that we do just for the good of man. It's not actually a big money maker or anything.
>
> But, eventually we saw it as a shortcut to just get a bunch of data online, and they just have bots that go and kind of scrape the data (from government websites) and then convert it to the standard. And we're just ingesting as much of it. We're starting with high population areas and getting as much as we can online. (PRI-02)

As we revisit this story through the lens of dependence, these observations are revealing. First, we understand this is not really a money maker for the company. However, second, it turns out that one dependency, a natural one in business, is between the online directory and its customers, in this case, the restaurants. The restaurants are paying to advertise, and even accusing the company of making their ranking "pay-to-play" (as cited above). This introduces another concept, left unstated in the vigorous – and certainly authentic-sounding – case made for the altruistic purpose of adding the health rating. It is a factor independent of the consumer reviews on the website, with the imprimatur of government adding legitimacy to it. But, from the position of a restaurant that chose not to invest in advertising, leaving its entry unclaimed, learning that the health rating was now part of the entry seems like it would provide an incentive to pay closer attention ("businesses are obsessed by their pages"), increasing the "stickiness" of the site for its paying customers and, to the degree paying more would help, increase revenue by greater participation.

So, in any event, there is a dependence on government data, however tenuous, as well. But the burden of managing this dependence, grew tiresome and the national company then came to depend on another business to collect the data whose business model was to use computer programs to sift through and harvest data from existing state and municipality websites that post their inspection data and then convert it into something useful for their online application. Now, we turn to that company, another link "downstream" in the chain of dependence.

Harvesting Restaurant Inspections: The Commercial Transformation of Public Data and the Market. In this closing scenario, we learn more about the company that crawled the restaurant inspection websites as part of their business model. However, in these interviews, we also see other concepts introduced that relate to the private use of public data and interaction with government that will come up in later cases. First is the idea that models and approaches may be different across companies using the same data and some may sit upstream / downstream from each other in the chain of use and dependence. So mapping interests, influence, uses, and effects can be complex. Second, one can see that the private sector may, through aggregation and its greater reach, come to "own" the public perception of what government data is and means. They can shape and add to the data after the fact, creating their own ratings, in this case, for the health practices of restaurants, but backed by the imprimatur of source data from government inspections. And, these indicators - essentially the most commonly available public information on the topic – are, in turn, likely to be shaped by the market. If the return on investment of acquiring the inspections covering the last 10% of the country's population isn't there, then those results won't be available in the app, even if they are made available otherwise by government. You see also that the restaurants understand the role these companies are playing, complaining to them about scores, rather than the government because they are the ones creating them. Finally, we learn something important about dependence: Companies can become dependent on government data even without government participation – all they need is some form of access.

The company that the national company turned to was formed at about the same time as the national company rolled out their standard. They didn’t work together - and took a different approach to doing something similar (PRI-04). A person employed there at the time explains:

> They [the national company] said “We are the de facto standard and everyone should come to us.” …Our standard allowed for more data. Their standard initially had flaws, like it couldn’t handle multiple inspections in the same day. So, the biggest difference between [their standard and ours] is they were essentially closed and we were essentially open. Which allowed us to aggregate more data, far deeper and easily than they did. *They had to get legislation and jurisdictional buy-in. We had to find a data source* [emphasis added]. (PRI-04)

Another person working at the company at that time added that part of the difficulty he saw in the national company’s rollout of the standard was that it didn’t accommodate the data being produced very well (which jibes with the story of the company developing it with input from only one other city):

> Then, the other really big thing is that the things like the result of the inspection and other fields, like violation descriptions, everything, usually have to conform to some kind of pull-down list…. It really it doesn't match schemas that you see across the country for how they're presenting the data, and it really doesn't match the naming conventions of severities and results that are in the food code, the National Food Code… Maybe they worked with some jurisdiction people on it, but it really kind of looks like, and I can't tell you this for sure, I can only tell you this as what it looks like from my opinion, and I could be wrong, it looks like they just made it into a vacuum for what was easy for them to do on the web, to digest on the web. (PRI-05)

From a mechanical standpoint, the other employee discussed how they obtained the data, usually without working with government, although they were mindful of the “Terms of Use” on government websites:

> Each has their own data standards and Terms of Service on the website. So, if you look at most Health Department websites, they actually don't even ... they basically say, "We wash our hands of this data." Basically, represented “as is,” there is no warranty. I mean, we actually had to follow some of the local laws. Like, in [one state], we had to sit down with the Public Information Officer before we started working with [that] data because there was a law saying you can't use the open data and sell it into mailing lists. [So], we read the terms of service. (PRI-04)

> [Otherwise] we weren't really dealing with [the government]. Our technology could handle OCR (optical character recognition) technology, like we did scrape an image of a

> pdf and get all the data and the correct images. [If the data weren't online] we actually didn't FOIA that because we just said, "It's not currently available." Because we had 1750 jurisdictions, out of what, 3200 jurisdictions - meaning basically a county (Note: Which, as we will see later, is purported to represent over 90% of the U.S. population). (PRI-04)

And, as the first employee noted above continued, working with government was not easy:

> We've tried, and it's pretty tough to get people to get people to pick up the phone. The administrators don't have time or interest. Anytime we try to talk to them, or we used to try to call and talk, and we've been hung up on. They think we're trying to sell something, and we're not. We don't even get a chance to explain what we're doing. Maybe we're not the best at that, but, so, yeah, just the effort [is a barrier]. Some jurisdictions it takes us, really, it could take us up to days or a week or longer or whatever to write and validate a bot, make sure it's doing what it's doing accurately, and that's a lot shorter than just even trying to get somebody on the phone, let alone work with them. (PRI-05)
>
> We probably have, like, 800 bots. The way we write this is there are bots that go out and scrape the data from the local jurisdictions. We take app's word for it (the government website or application that displays its restaurant inspection results). We said these jurisdictions just move too slow, and this is going to take a long time, so let us collect the data and then go. (PRI-05)

As he then observed, the logic really just dealt with return-on-investment and the size of the population being covered by the data:

> So, New York City is going to be more important than Omaha. You know, state of Florida's going to be more important than Omaha. And like Denver, the city of Denver, is probably more important than the state of Nebraska and Omaha. I mean, if you look at the population, last time I looked when I was running it, we had like 93% of the US population. (PRI-05)

This company, in addition to transforming the data for the national company, developed their own standard, or "algorithm score" that is now presented on the national company's app and website, as well as through their own stand-alone app. It attempts to "correct" for some of what they see as problems in the current delivery of information about inspections. Again, one of the employees explains what they are trying to do, along with the problem with standards in this area generally:

> That company's national standard or any other standard that's based on just reporting with the jurisdiction reports doesn't allow you to really do comparison to comparisons. For example, what ends up happening is, and let's pick on California because they've got very discrete, they've got dense population, with restaurants that literally are across the street that could be in different health jurisdictions.
>
> And so what ends up happening is, and we see in feedback from [the national company], is that one place has an A, one place has a different score, and what the jurisdiction's reporting and how hard the jurisdiction and local laws reflect how challenging it is to get that grade, you'll have restaurants complaining to us and say, "Well, I got an A," and of course we'll always tell them well that goes back to your health jurisdiction, but, "The restaurant down the road got an A, and they had three criticals. I got a B, and I only have one," and it really just comes down to how the jurisdiction does their grading now. That's an extreme case. I'm just making that up to illustrate the point.
>
> Now, our score, like if we give you a 36, you're a 36 in your health jurisdiction. But after that, but that can be used to compare to a 36 in this jurisdiction or a 36 in another jurisdiction, is the 36th percentile worst restaurant in the area gives you the ability to have a cross-comparison that kind of averages out. Our score takes into account when it does its calculations the local deviations, your local jurisdictional severities, and then the score becomes kind of the way to do a comparison-to-comparison. (PRI-05)

The other employee continued the story:

> We have come up with a historical based scoring system that lets you know how a restaurant really is clean compared to its peers, and based on its history. Because they get re-inspected, and the re-inspection becomes the new standard of record, but it doesn't speak at all to how many times they were re-inspected, how many times they failed prior, what really is their historical performance been. We consider ourselves like the credit-rating agency type restaurant inspection. Your credit score doesn't reflect the last time you paid bills. It reflects how long you've been doing it for the last two years. That's really what we're trying to bring to market. That's really for us the focal point. We have a paid app now, just to try to make some money.
>
> *It just really just comes down to us trying to be the TransUnion or Experian of restaurants. Right? We want a way to kind of give a consistent number to "your FICO score is this," "Your [company] score is this," this percentage. Right?* [emphasis added] (PRI-04)

In discussing the barriers to adoption of a standard, they, too, referenced the resistance they encountered from restaurants and the restaurant industry:

> Yeah, typically we do get a lot of feedback, daily, from restaurants. Never, rarely positive. We actually did at one time try and reach out to try to integrate. There's three kind of customer interactions, there's the restaurant associations, and there's various of

> them, there's the chain restaurant associations, and we've gone and met with them, we've sent people down to give presentations, and more often than not they're not really receptive. They do, and these people care about being clean, they don't want to ... No one's going to tell you to your face that they don't want ...They're not going to say "Well, I don't want this information out there.
>
> You know, but it all comes down to that. We get a lot, generally we get a lot of feedback from restaurants. Most of the time, they're pissed that we rate them at a certain rating. You know, "How dare you come up with your own rating, the jurisdiction gave me this, why do you do this?" And I say, "Well, it's our opinion." We're public with our equation. It's on the website. Even the credit rating agencies don't give you their equations, right? (PRI-05)

Finally, it is important to note that there are yet other models using this data, focusing on different markets. That is, the ecosystem is yet larger, with opportunities for yet more logics of dependence and influence to be mapped and investigated. Both employees referred to other opportunities and models, some that even they had engaged in:

> Well, but there are some certain chain restaurant owners who have come to us and said I'm not happy that you made this public, but I want you to help us keep an eye on our restaurants...because we find out way too late that someone's had a bad inspection.
>
> Now we don't generally play in that business as much. We're trying to be more consumer focused. There's another company and they're focused purely on the B2B. They have customers that are big chain restaurants, and they try to provide this data to them. They would appreciate a standard, as well, because they do the same thing we do, they scrape data.
>
> So they basically went to locations with 50 to 100 or greater of restaurants and said, "You have establishments in these ten jurisdictions, or 20 or 50 or 100 jurisdictions. We can tell you what their health department inspection scores are in a single spot. So, we can show you where your liabilities are so you can send your trainers in. You can send your clean-up crews. You can do that." (PRI-04)

Another relevant example was one that their own company had pursued, outside the consumer market. The other employee related:

> We sold to insurance data brokers who were basically trying to build models to sell to insurance companies and say, because a site survey was done by the sales rep. When you do a commercial restaurant policy - and I don't know about you, but a sales rep isn't a licensed health inspector and probably doesn't know what to look for. So, there's a conflict of interest there, so ... And also, this report comes out at least annually, but if not,

> semi-annually or quarterly depending on risk. So, getting fresh information is a better judgment of rate or risk than a site survey done by a sales rep - once. You know, so there are different industries who look at this data differently. (PRI-05)

All taken, even with very limited support from government, the private sector succeeded in finding a way to use, and even produce revenue from, restaurant inspection data without the participation of government, but not without trying to engage, or even browbeat them. However, in the next case, we will find that government is more receptive to industry overtures, as they see a chance to benefit themselves from the arrangement.

Genealogical Data

In this section, we look at mutual dependence in a specific phase of the information lifecycle: The decision about whether to retain government information and for how long, and the decision about how access will be provided to it once it has reached the archive. These decisions are placed at the center of the analysis by looking at the impacts of dependence on "no-cost" agreements made by government archives with a private sector company for services to preserve paper records by digitizing them. In turn, the company is granted a temporary, but exclusive right to sell access to the digitized copies. First, we examine the impact on <u>access</u> to data and, second, the potential impact on decisions about <u>which government data to retain</u>.

Decisions about the Retention of Government Information: A Precursor to Access

Private sector influence at the archive is relevant here due to a fundamental conflict that sits at the center of the life of information. Access cannot be granted to documentation that has been destroyed or discarded. Some hypothetical examples: That expired contract you wanted to see – sorry, the retention schedule allowed us to destroy it two years ago. Documentation from meetings about coal plant emissions? Gone. Waiting list for low income heating assistance – how many people never received aid? Sorry, the government-approved retention period was

"Keep until no longer useful, then destroy" and it was destroyed by the previous gubernatorial administration when they terminated the program. But, in turn, keeping copies of all the information coming in and out of government is not fiscally or technically feasible - nor perhaps even desirable (e.g. "the right to be forgotten"; Jones, 2018). While there are established professional frameworks for appraising the value and determining the appropriate retention period for data, the literature suggests that only about five percent of all government records are determined to have enduring value (Bradsher, 1989).

Records Scheduling and Appraisal in the States. The following excerpt is taken from field notes made while observing a committee that serves as part of a state appraisal and records-scheduling process:

> The electronic records oversight group had been meeting for more than an hour to review plans presented by state agencies for how new and existing computer systems would maintain government records. As explained on the related website, this process has been put in place "to ensure that agencies effectively manage electronic records with long-term value (10+ year retention)." The next item on the agenda is the historical organization's own visitor registration database, an electronic listing of individuals that had registered to view materials in the reading room. Early in the discussion, one of the members of the committee noted that the records were scheduled to be retained for 10 calendar years and asked: "Why not nine years, 364 days?" (Field notes, n.d.)

In state government, all fifty states make use of some form of "records retention and disposition schedules" that identify sets of records and how long they are to be kept (COSA, 2017). Similar approaches are used at both the local and national level. In one state, this "records scheduling authority" resides with a statutory body made up of agency representatives who, with the force of law, determine the length of time the records must be retained and what is to be done with them at the end of their useful life. Their decisions are based on proposed retention periods that are developed through an appraisal process that may involve input from agency employees and records managers, archivists, elected and appointed officials, along with other stakeholders. The question remains, however, which current and potential future users - and uses – are considered in the deliberation and how the eventual ability for citizens to access the data that is saved comes into play. After a discussion of the professional considerations used to make that determination, we will examine how private sector use of archival data comes in to play.

Applying Appraisal Theory. Much of the work performed in determining this value in state government today has its roots in guidance developed by T.R. Schellenberg (1903-1970), a former assistant archivist of the United States. Schellenberg's framework divides the values applicable to records into two main categories. *Primary value* is considered the "value to the originating agency." This category contains those records of legal, administrative, or fiscal value to the agency – basically their usefulness in fulfilling the requirements of their daily operations. *Secondary value* is defined as the value of the record to "other agencies and to non-government users." This category is further decomposed into two types: *evidential value* and *informational value*. Evidential value consists of "[e]vidence public records contain of the functioning of the government body that produced them." Informational value refers to the "[i]nformation that is in public records on persona, places, subjects, and the like with which public agencies deal." (Schellenberg, 1956).

While this framework provides general guidance for identifying what materials are of "enduring value" and thus suitable for permanent retention in the archives, the retention of the rest is mostly left for resolution with government agencies based on its primary value and useful life for their operations. The process does involve discretion, however, and is implemented in an environment where more information is being created every day, technology is changing, budgets are tight, and space and staff resources are limited – as is the tenure of employees involved in the process. Thus, such decisions are by necessity imperfect and may even be revisited as the view of the value of information – or the cost-benefit calculus of its ongoing upkeep and storage – changes over time.

Beginning at the End: Making Archival Data Available. Perhaps hidden in the discussion so far are some aspects of access to government information that are not intuitive. First, in today's world, and in what's left over from the past, information exists in both paper and electronic form. The older the information, of course, the more likely it exists only on paper. To the extent that the *record copy* (official copy) of data exists only on paper, the practical costs of access involve scanning or duplication (creating a digital "surrogate"), and the technical and human resources needed to perform these actions to make it available. Second, during its useful life at a government agency, it is likely that a small part of the information they create and receive is made available to the public. That is, while some access is restricted by law (say, to working papers in legal decisions, or personal information), the rest may exist in agency systems or file cabinets, and is only made available via Freedom of Information Act or state open records requests, not on an agency website. While some agencies and programs may be more transparent than others, it seems very likely that there are few if any cases where all data held by the agency that is open to the public is published on the web, if only due to the practical resource and management constraints on doing so.

This leads us to the importance of decisions about the retention period for government information. For records whose retention period is deemed to be permanent, their ultimate destination is a state archive, where the information is consolidated in one location, and whose primary roles are preservation and providing access. For those records whose useful life is anything shorter, access is left to what an agency can or is willing to provide online, supplemented by responses to individual requests by the public for copies for those records that are open. As a practical matter, then, decisions about how long data is retained and how access is provided by the archive represent a fulcrum for determining the information's availability for use in government transparency, accountability, and other critical purposes outlined above, including the kind of secondary uses by commercial parties that we will consider in this and subsequent cases.

The stakes are high. In the state referenced in this case, the archivist described the current holdings of paper records deemed to have permanent retention as approximately 49,000 cubic feet or, as he put it "…a professional basketball court stacked to the rim with paper - the equivalent of about 5 million sheets of paper" (PRI-06). Nationwide, a 2016 survey of the states found that they held over 2,200,000 cubic feet of paper records from state and local government, with median growth of around 50,000 cubic feet per year (COSA, 2017). While this establishes the size of the corpus of information involved, generally, another result from the survey confirms the relevance of the issue at hand, the potential for commercial influence on access. With 44 states (and/or territories) responding to the question "How do you make actual records or information from those records available via the Internet?" just over 77% report "Access provided via vendor websites (e.g. Ancestry, FamilySearch)" (COSA, 2017).

A Paradigm for Preservation and Access to Archival Data: The Digitization Partnership

Providing access to all this paper is an almost insurmountable challenge. In the same 2016 survey, the median funding devoted to archives and records management of the 44 states and/or territories reporting was just .007% of each states' total state government expenditures for FY2016. Nineteen states (45%) reported building, restoring, or revising staffing as one of the three top issues / initiatives they faced, with space planning and management in the top three for sixteen of them (COSA, 2017).

With that in mind, digitization of paper records is an expensive and time-consuming process that involves staff expertise and oversight. The level of effort and expense varies with the age and form of the records, and includes such expenses as consulting and evaluation of appropriate conservation methods, preparation of the material to be scanned, rental or procurement of the equipment needed to scan the records and personnel to run it, storage and movement of materials, preparation of metadata (data about the data, prepared according to professional standards for access and indexing), quality assurance testing (a page misaligned in the auto feeder for example), re-assembly of documents for refiling, and related tasks (FRMC, 2019).

In the face of these challenges, however, archives have been approached by both non-profit and commercial interests who are willing to "help." The public has long been interested in the use of government records for genealogical research - birth, marriage, military, or other records containing personal information. This interest exists across society as a whole, but, for religious reasons, some of the first efforts to engage government en masse in expanding access to these records originated in the Church of Jesus Christ of Latter-day Saints (LDS). These efforts began with microfilm, pre-dating today's modern digitization and access methods. The Genealogical Society of Utah played a principal role in this outreach. An employee in charge of such partnerships for a government archive recounts his experience with the history of this process:

> We've had different but similar relationships prior to the digital era with organizations like the Genealogical Society of Utah (GSU). We've had long-standing agreements with them to help preserve and provide access to public records particularly at the county level through a partnership in which we would assist in identifying, arranging and preparing

> collections to be microfilmed. Then GSU would get volunteers in to actually do the microfilming and they would get copies of the film to go in their Granite Mountain Vault (in Utah) and we would get copies to add to our collection. So, we have participated in these kinds of agreements for a long time prior to the digital world. (PRI-07)

While government archives vary in the scope of their responsibilities, this archive is also responsible for county records schedules. He continued:

> So, we played a role in trying to help preserve county records and we used our partnership with GSU to do that. So that was the best way for us to not only to help preserve public records in the counties, but to get access to it as well so we can provide our patrons access to it. So that's really the primary way that we've accessed that information and that our patrons gain access to information is through that microfilm created through the GSU partnership. (PRI-07)

As digital technologies developed and the Internet became widely available, there was a natural migration to their use, both by GSU and by commercial firms, like Ancestry.com, a large online consumer-focused genealogy website. The non-profit origins and religious ethos of these efforts are relevant as we begin to look at mutual dependence in these partnerships. Having worked directly with them in his home state of Utah and at the archives in the Midwestern state for many years, the archivist shared his perspective on the story:

> GSU and what became Family Search, they're basically the same thing, just changed names. In the past on projects, they've been very willing to devote their time and resources to help preserve records that to some extent, don't have genealogical value or that even records that we would not allow them to retain copies of or access. Because they wanted to show that they were interested in supporting preservation of public records, outside of their own interests… I haven't really experienced that with any other vendor or group that we've worked with. So, they have helped us with microfilm projects and with digital projects in which they weren't directly benefiting. (PRI-07)

In summary, then, there is a history of partnership between external organizations and the state archives, where these organizations, in this case, GSU, would donate labor and use of equipment (although, to complicate things, for at least some records, this involves prison labor) (Bauer, 2015) and the state would prepare the materials for microfilming. So, we see a mutual dependence develop where both parties appear to benefit. And, as the archivist mentions, the non-profit organization exhibits some altruistic behavior, occasionally agreeing to help with projects from which they do not appear to gain direct benefit. Yet there is an aspect that will become pronounced as our examination continues, and that is the emphasis on genealogy that informs these efforts. To the degree that archives depend on these types of agreements for both preservation and delivering online access, the question occurs as to how that might bias the information that is preserved and available overall toward content with value for particular audiences. The private sector naturally plays a part in amplifying their demand to the potential sources of such data, which, in a resource-constrained environment may result in neglecting records of interest or import to other constituencies and users. In this way, as I will pose later, it is possible that market values may work their way into not just decisions about what to make available, but about what to retain. In practical terms, as far as the (online) access in these examples, it appears they already have.

The Embargo. We take up the story with a for-profit company mentioned earlier, Ancestry.com. Here, the agreements entered into are much different than they were with GSU. They depend on what they call an "embargo," that is, a restriction by the archives on public access to the material that is digitized by the company for multiple years so that Ancestry may sell access to the public during this period without competition. From a copy of the state agreement provided by the archivist (dated 2009):

> [T]he licensed materials may not be distributed or resold by [archive] to any company or institution for any purpose, and are solely for the use of individual patrons. [Company] grants [archive] rights to post the images online 5 years from the date that [company] has posted the images.

In other words, the archive agrees to limit access to the public to the newly digitized records, preserving the commercial value for Ancestry, in return for the digitization of the material, helping to ensure its long-term preservation and, consequently, potential for later access. In the words of the government employee who negotiates these agreements:

> Generally, they're asking for access to the materials and for a license to sell the materials with an unrestricted license. Because if they're going provide all the resources that it takes to digitize, create the metadata, and publish it online, they don't want us to be able to revoke it or limit their use of it and so forth.
>
> In exchange, we ask for a digital copy in a very specific format that we can use and we're able to preserve. So, we have standards that we expect them to meet and in terms of quality of those files and the type of formats that they're using. In our case, we ask for - and on a specific timetable - access to those materials in real time. So as soon as they hit their website, we want our patrons - essentially our state's residents - to be able to access those immediately on their site at no charge.
>
> So, we're getting a digital copy in a preservable and usable format and we're getting immediate access for our core audience. Then after, let's say, a three- or a five-year embargo period that allows them kind of some exclusive period where they can recoup their investment, we ask to be able to open up that material to make it freely available to everybody worldwide. So, they have a very limited time to sell that exclusively and recoup their investment before it becomes free to everybody. (PRI-07)

Here we see a clear example of mutual dependence and, as a result, digital access that could be made broadly available, as would be the case with FamilySearch, the non-profit successor to GSU, that is instead delayed to prop up the commercial value of the records for Ancestry to generate a return on their investment. While it may be for a time certain, it does create two classes of access to government information based on residency. The calculus is revealed in further discussion:

> Well, I think the way I look at it, online publication is really - at this point, we still consider it special access. It's not mandated by law. It's doesn't really come under, like we're providing free access. If you want to come here and look at it, we're providing free access to it. We're not restricting access to it. (PRI-07)

Let's recap the logic. Prior to the agreement, no one has digital access to the material. On the other hand, given the time and sometimes money (distance) required to travel to the archive and view the material, as a practical matter few may really have access to it at all. And, as a paper record, it requires the continuing cost of physical space to maintain it and a temperature-controlled environment to preserve it, while it naturally decays a bit each year. With the agreement in place, the material is digitized for free to standards that can begin to ensure its long-term preservation and access, and they negotiate to at least grant access in the short term to the same citizens who pay their salary. And, eventually, the data / records can be made available to the public at large and they can do what they want with their copies. But, it's a little more complicated than that, as the employee continues:

> After the embargo period, then most of the restrictions on those materials drop and we can publish them online for free and give access, open up access to those materials for anybody. What we can't do, usually there's still two restrictions that apply. We can't sell them, and we can't give them away in bulk. We can't give all that data to another institution or another vendor because obviously it's in their (Ancestry's) interest to restrict competition.
>
> So that's kind of the standard agreement and that's how it works. *After the embargo period, if we can put them online, we can open them up to the world. The problem is in most cases, we don't have a website that we can easily deliver that content on. We have an online archive, but it's not robust enough and the architecture is not such that we can easily accommodate and provide easy access to that data* [emphasis added].
>
> In most cases, like with Ancestry, the embargo period we have with Ancestry, *everything that we've given them is past the embargo period, but we're not delivering access to any of that content* [emphasis added]. So even though we can make it freely available to the world, we're not, just because of our own limitations. *I think companies like Ancestry, when they make these agreements, they know that most of the archives they're making agreements with probably aren't going to serve (make available online) this stuff* [emphasis added]**.** (PRI-07)

The idea that access matters and that various forces may shape it, intentionally or not, is not unique to this research. But, here we have a clear illustration reinforcing the premise of our study of influence and mutual influence in commercial use of government data. Private sector interests may determine priorities for digitization and the practice terms of access to those records. The overall framing of what is at stake here is provided quite concisely in the literature of archival science. For all intents and purposes, what's online may become "what exists":

> Decisions about which records to describe in greater detail, and which to digitize for remote access, will influence the characteristics of the documentary past for many users of archives. Materials that are discoverable and accessible remotely will enjoy more use than their physical counterparts, because remote access removes barriers of distance and time. If remote access becomes the predominant way in which most users discover archives and interact with their contents, then the on-line collection becomes *the* collection for many users. Archival exhibits and online collections are highly-mediated creations that are influenced by funders or sponsors' interests and by archivists' views of what is valuable or interesting.
>
> Selection of what goes up on the web privileges a tiny portion of the archives, chosen from a larger body of archival material which itself is only a small percentage of the documents that once existed. Yet archival exhibits and on-line collections provide few clues about the basis for selection or the existence of related physical and digital materials. Taken too far, this strategy can produce superficial digital collections, removed from their original provenance and context, that reinforce dominant master narratives of progress, nationalism, ethnic superiority, patriarchy, technological determinism, or whatever those making decisions about what to digitize decide to emphasize. (Hedstrom, 2002, pp. 40-41)

Tracing Back Access to Appraisal: Examining the Potential Influence of the Market on Decision Making about Retention

Now that we have considered the potential impacts on access of mutual dependence with commercial parties to digitize paper records, the quotation above leads us to a question. With the rise of commercial intermediaries focused on obtaining revenue from particular domains of information, the value of the secondary use for these records is empirically established. However, as discussed above in the section about the appraisal process for government records, decisions are made about length of retention by considering, among other things, *their secondary value*. In the case where budgets are very tight, and, among them, the cost of preservation and the constraints of available space are unavoidable factors, is it possible that the cost savings and associated opportunities for access introduced by these digital partnerships might affect decisions about what to retain?

The argument being made here is not one of direct influence on appraisal by the private sector. Instead, comprehensive universal electronic access to government information is a laudable goal but made a chimera in this situation by real-world constraints. Once it is established that a particular type of record has a ready, known audience and a private sector subsidy for preserving it, however, it seems possible that the secondary value may somehow be seen differently or prioritized over other records. Bluntly put, as a particular use or audience is amplified by commercial interests in one area but not another, perhaps favoring certain socioeconomic groups or majorities over others, could the values associated with these market perspectives somehow seep beyond influencing access into decision making in appraisal and retention, with consequent impacts for the shape of our picture of the past, and understanding of and accountability for government actions? Or, perhaps less severe, but with a similar outcome, is it simply that some sets of records will become much easier to access, with essentially the same effect?

These are large questions, far too large to be substantively addressed in only a portion of one chapter of this research. However, to illustrate the situation in which such forces play out and the malleable dimensions of the context in which appraisal and retention decisions are made, the remainder of this section takes an ethnographic approach to one episode of the appraisal process. It uses excerpts from a series of field notes to follow the journey of just one type of record, that of the pharmacy assistant, for a government agency as a decision about its retention and disposition is made. The story focuses on records of an agency that regulates this occupation, as well as investigates complaints.

Establishing the Value of Government Records: Access and Pharmacy Assistant Records

I began by tagging along with the archives staff responsible for records appraisal and scheduling to a meeting with the pharmacy agency to walk through a jointly planned update to their existing records schedule (all names are changed to preserve confidentiality). I then followed the fruits of this work through the major steps in the process, including a detour to a local professional conference of archivists where I talked with a representative of Ancestry about how they assessed the commercial value of records. The story ends with a meeting of the state oversight board where they engage the issue of their historical and genealogical value. By doing so, I hope to give the reader a better sense of the subjectivity and "open" nature of the appraisal process, how altruism is set against practicality, and provide some sense of how the tradeoffs brought into play by commercial interests in government records, in this case for genealogical purposes, might grow to have larger influence in the process over time.

Making Decisions about Government Records: Influences and Outcomes.

> I went with Mark, Mary, and their database administrator Carla to meet with Judy and Marvin, their contacts for the agency records, at their office downtown. The meeting was held in a conference room tucked around the corner from the entrance. Judy began the meeting by handing out copies of their records schedule with proposed changes noted next to some of the entries. There were fifteen entries on the schedule, with changes proposed for five or six. (Field Notes, n.d.)

The ensuing discussion provides some insights into how records are appraised and how the conceptualization of enduring value and the role of access changes over time with changes in underlying record keeping and access technologies. While several situations discussed do not bear directly on the idea of mutual dependence and influence that we are focused on here, this brief recounting of the visit will provide context for the process of appraisal that will be needed in later discussion of how influence might be introduced:

The complaint files include material about the investigation and prosecution of complaints against pharmacists. The retention period on the current schedule said "Retain five years then transfer to the archives." Judy proposed that the retention be changed to destroy the records at the end of five years. Mary agreed with that, as "they tend to be so highly confidential." If approved by the government records oversight committee, this would allow for the destruction of older records held at the agency offices. However, while not mentioned during the conversation, I knew from my review of the archives database that approval of this change would also result in the deaccessioning and return to the agency for destruction of 38 cubic feet of complaint files that began with records from over 40 years ago.

Mary then asked "Is five (years) OK? - do you need to keep them longer?" Marvin agreed that this was a good question and said they might look at that. The subject seemed pretty much settled, although Mark (the government archivist) had brought an iPad to the meeting and was beginning to look at a set of the files that was available through the agency website. The schedule for the series said there were no restrictions, yet Mary had mentioned earlier that they were mostly confidential. Mark pointed out:

"Yeah, that omission is an artifact of how old the schedule is. That wasn't part of what we thought much about back then. We haven't gone back to update these."

To me, it sounds like confidentiality constrains value because of its limits on access, so this made me wonder about all the records that may be sitting in the archives now with this same issue that have not (yet) been reevaluated. Mark continued to look through the records that were available online, he noted that they have been "making a switch" regarding complaint files. "They used to almost all be archival, but it turns out a lot of them are confidential, so we are now moving to consider them to be transactional and not of enduring value."

Mark has continued to look at the complaint files on the website during this discussion. At a break in this conversation he says "I'm still obsessed with this stuff that's on the website – You're doing this because of law?" Marvin responded, that yes, although these were just the orders, there was more confidential supporting material in the files. "It won't contain all the complaints we've ever received – Like, we don't keep track of unsubstantiated complaints. 'I got the wrong med in a bottle' – I got one of those last week."

Mark continued: "I don't see why we wouldn't want these…You kind of get the sense of what a pharmacist did – some of them are pretty detailed. "It almost feels like it's a separate series - (to Mary) see where I'm going? We should reflect your practices and intent here – if you have summaries (from that long ago)..."

The archivists left it that they would likely want to archive the records on the website, and then perhaps set up some way to "harvest" them in the future, just keeping the orders, but not the supporting material that was confidential. What I found interesting here is the idea that, as an overarching concept, their stated approach was not to keep "transactional records." Yet, on further examination of the records - enabled because of electronic

> access in the meeting - they saw them to have enduring value. I was also left with the impression that it was the more prurient aspects of the records, explaining the negligent behavior of a pharmacist in some detail, mixed in with "get(ting) the sense of what a pharmacist did," coupled with their lack of restrictions - and that they already existed in electronic form - that gave them their value.
>
> (Field Notes, n.d.)

The issue of retention and value came up again in a long conversation about how records of pharmacist and pharmacy licenses and associated cancellations would be addressed. While they were permanent, with the move to an electronic system two years ago, the agency was interested in not having to keep the paper that was being scanned into the system for new applications forever.

> Mark felt there would be a way to receive electronic copies from the new system into the archives periodically, and then allow the oversight committee to destroy those records that had been scanned at the end of the 10-year useful life that they had recommended. I asked about the cancelled pharmacist license and what the basis was for holding on to them for 10 years. Marvin thought there was a statute, Judy wasn't sure, so they both suggested they could ask the head of the agency, who had been there 10 years. Marvin went to check. When he came back, he said "There's nothing in the statute that she knows of about retention, it's just a practice." At this point, we were close to the end of our allotted time as lunch was approaching. There was one type of document left - pharmacy assistant registrations.
>
> Marvin questioned their value – "You wouldn't want those would you? He said "There's a massive amount of them and the turnover is staggering. Anyone can be a tech – many of them have criminal records...it's a different slice of society."
>
> Mark and Mary seemed to agree, but Carla – the database administrator - who had remained silent until now, interjected after the term "slice of society": "Well, it does capture a group of people that might not be captured in another source." Mark joked "We're not going to bring her along anymore" and we all laughed. With that, it was resolved not to change the schedule entry and that the archives didn't want to take those records.
>
> (Field Notes, n.d.)

While in no way a comprehensive analysis of the appraisal process for government records, this brief excerpt demonstrates how fluid and subjective the process can be and the factors that can go into the decision: The revisiting of past decisions through the light of changes in philosophy, personnel, and technology, and differing conceptualizations of value and its relation to later use. It seems then that the idea of secondary use can definitely play a role in the decision.

To complete our analysis of how agreements with commercial parties to exchange access rights for digitization services might come to impact the appraisal process, I will focus on the value of the pharmacy assistant records – the issue raised in the last exchange during the meeting - as they later move into the final stage of the appraisal process. But first, a detour to learn more about a commercial firm's perspective on value and how they might view this particular record series.

In the run up to the records oversight committee meeting, I had the opportunity to sit in as an observer at a booth at a local archives conference. The field notes pick up from there:

> Mark (the state archivist) spotted a representative from Ancestry.com, an industry-leading genealogy website, at another booth and called him over, saying "You'll want to meet this guy."
>
> Brad is a manager for the company and apparently knows the archivist well. Immediately after we exchanged introductions, he asked "When is your state going to get their vital stats records online?" I asked in return "Why do you think they're valuable?" then went on to tell him about the purpose of my study and asked him to talk about how he – and Ancestry.com – determined the value of records. As we sat down, Brad knelt on one knee on the other side of the table from us so that he was at eye level and we began a conversation:
>
> "Ancestry is "name-centric," so that's why the birth certificates and vital statistics records are valuable. Once you have those, you can then extend the "story" to employment or census records."
>
> We talked more about name-centric records. He brought up the high value of Pullman Porter (railroad) records. "They were the largest employer of black people in America. So, the odds are that if you were African-American, one of your relatives may have worked there at one time."
>
> Then things got even more interesting. I asked about state records and Brad said: "Well, an example might be the cosmetologist licensing records we are digitizing in California right now. They cover the early 1900's through the 1940's and are very diverse." I seized on his use of the word "diverse" and stopped him, asking him to explain further. He said they "documented diverse populations that might not otherwise

have been documented." The gist of our conversation was that "there weren't a lot of sources of records of minority populations during these time periods, so a set like this might be one of the few places where someone could be identified – their occupation, where they lived, which might be able to be tied to other records."

He gave another example: "I was talking with the archivist in Montana about teacher records. In this case, they turn out to be predominantly single women in Montana, and this was before they had the right to vote. So, they're an example of a population that's not that well-documented and could have value for that reason."

Mark and I both remembered Carla's observation about the possible value of the pharmacy assistant records and I asked Brad if he would be interested in them. He answered:

"Well, comprehensiveness is important. They would have to document an entire jurisdiction, not like just one county. And, also the degree to which they are relational – like tying parent to child, husband to wife –that's why census records are one of the best. …What's the story that the record tells? If it just has name on it, that's one thing, but if it has other information – say, the value of the house they owned, then maybe the value of the houses around it, you can say "Hey, they owned the nicest house in the neighborhood."

Mark (the archivist) mused to both of us about this:

"I hadn't really been wearing my Ancestry hat when we were appraising them. It is somewhat challenging to consider the value of records with Ancestry and genealogists in mind…It's also difficult to invest in keeping contemporary agency records that would not have value until 100 years from now…I don't even think about that in a systematic way – it's not our job. Our job is to document the functioning of our state government and protect legal rights."

(Field Notes, n.d.)

It seems clear from other conversations we've had that one of the major users of the state archives is genealogists. This means that the observations by a proxy in the form of Ancestry.com are an important insight into values the public might place on records. By definition, records transferred to the archives no longer have a "useful life" to the state agency that created them. And, one of the Schellenberg's appraisal criterions is "informational value." Yet, in the conversation at the pharmacy agency, Mark had talked about the move away from keeping transactional records. The demand for records by genealogists brings into focus this idea of a time lag as it relates to value and highlights the dichotomy between the ever-present opportunity for reappraisal based on current perspectives and the potential for different valuations off lurking in the future. One can see from the examples provided by the Ancestry representative, however, that the longer-term value of a record was seen through a unique lens that may not always be complementary to the principals of appraisal used by an archivist. And, in turn, they have resources to enable preservation and access, and, by definition, their business model serves to proxy public interest, albeit in what is essentially a neoliberal, market-based way. It seems they would not offer to digitize records of interest to populations from which they cannot generate a return on the investment.

The Final Decision: Subject to Change. At the end of my time with the archivists, I attended a meeting of the oversight committee where the state agency records were on the agenda, along with several other state agency record series. By the time these records came up for discussion, many of the attendees were gone, but the members of the committee, representing the state legal counsel (Linda), the central services agency (Jeff), a government library (Ben), and the archives (Mark and Pam, who is his supervisor) were all still in attendance. Judy was there as a representative of the agency.

> In reviewing the new schedules and the proposed revisions, Linda noted some incorrect legal citations on the restrictions, and the committee voted to table the changes on the sets of records that were now going to be folded into the new series for the licensing database. Afterwards, she asked the question: Is there some reason you picked the Pharmacist to transfer and nobody else? Mark responded "Ah, that's a very good discussion. We thought a lot about it." He then proceeded to recount our conversation with Brad from Ancestry.com, ending with the following observation:
>
> "As an oversight committee, we have to think about the value we assign to records. And, while family history and genealogy is one value, it may not be the way that the State Archives is normally approaching it, that's not the primary reason we keep records. So, if we have to records around for 60 to 70 years in the hope that the Ancestry.com of the future will make those available to people, is that enough justification for expending state

resources on the preservation of those materials?"

Judy spoke up. "I did bring - I don't know if you guys are interested - some of our older pharmacist records, (*there was sort of an "aww" sound from one of the members*), just so you can see what we'd actually be destroying - it almost seems...like the electronic copy doesn't do it justice." She passed around several small manila folders used to hold the registrations - a single sheet, tri-folded, with a photo stapled to it, including the individual's educational history and test scores on the pharmacist exam. After a bit of deliberation, they had the following exchange:

Mark: So, I kind of lean toward Mary's assessment. The resident pharmacists would have historical value…I'm on the fence on the pharmacy assistants. Basically, it tells you their name, address– it doesn't tell you where they worked. So, to me, that doesn't quite hit the threshold…

Ben: These are people who are going to slip through the cracks. People who are pharmacists – yeah, you're going to find them all over the place. And also the accessibility – a lot of the later records might be private and unable to be accessed, but a person could be found through these pharmacy assistant records.

Mark: I am very sympathetic to that – the issue then arises "Where does that stop?" Or does it stop? I mean, you can think of all the kinds of licensing records that are generated by the state. We have traditionally kind of cut it off at professionals. And, I've always been a little uncomfortable with that – we've talked about it for as long as I've been on the oversight committee …Well, we have another quarter to decide...

Pam: We keep telling him that appraisal is an inexact science. (*Speaking to me*) You should put that in your paper. They're inexact results because it's subject to change.
(Field Notes, n.d.)

As we have seen in each venue – the offices of the agency and the deliberations of the committee - the construction of the value of a record is achieved collaboratively, with the interests of some current (and all future) publics represented only theoretically. And, while it may take place within an appraisal framework known to state archivists, it plays out in a context of forces – changes in technology, resource limitations, privacy concerns, constraints imposed by the experience and views of those involved in the process, and bargaining about what is valuable, to whom, and for how long. All these changes, too, are set against a background of the ongoing loss of historical context that occurs through employee turnover and periodic reorganization.

For the archives to accession records, they must be deemed no longer useful for a business purpose. Rather than liberating the decision from the constraints of practicality, this increases the seeming arbitrariness of decisions about what is kept and what is not. The situation is, in turn, compounded by a process that allows for decisions to be revisited over time, resulting in records that were once permanent being destroyed, but without a complementary capability to recapture a past that was not preserved. The current transition from paper records to electronic format means that the paper records left behind may, for all intents and purposes, disappear – either through destruction, disintegration, or because they have become invisible to online users. In the case of the pharmacy records, both the agency staff and archivists see value in continuing to maintain records past their useful life, yet we know this is not always the case.

In the earlier part of this section, we saw the exchange of value between the public and private sector, with at least temporary limitations on access provided in return for services (digitization) that could extend the life of public records. The private sector depends on archives for the records they sell and for them to grant them the right to restrict access long enough for them to make money by doing so. In turn, the archive depends on the private sector to digitize information to extend its useful life and provide access. In the case of the appraisal, the impact is not so obvious, and perhaps just now emerging. We see that the mechanisms where value of government information is established can be sites of contestation and that the consideration of future secondary uses and value involves issues of equity and access, two dimensions that are wrapped up in the very commercial agreements that help paper records to be preserved and citizens to use them. But, the values of those funding the work and the records that end up being preserved as part of these agreements are being driven by a market model in which the equation may only partially align with the goals of a government function. In times of poor funding, perhaps something is better than nothing. But in this case, as resources brought by the private sector shape the future, they may end up shaping the past as well.

Driver History Records

In the last case, we saw a government agency trade exclusive rights to sell public records to a private sector company for services that helped them achieve their mission. In the case we look at here, Driver History Records, there seems to be little benefit to the mission of government, but a very large benefit to its bottom line. As noted in the introduction to this study, there are numerous cases where states receive large sums of revenue from sale of information they collect about drivers, primarily to the insurance industry for rate setting on auto policies. While these numbers may include a wider selection of records, in 2018, the State of Texas received just over $69 million in revenue from these sales (Texas Comptroller of Public Accounts, 2019), and from 2010 through 2015, the Commonwealth of Pennsylvania received $157 million (Blacher, 2016). On the private sector side, in the five year period 2005-2010, the company ChoicePoint, Inc. alone paid the Ohio Bureau of Motor Vehicles more than $36 million for data about drivers (Guillen, 2010). Unlike the two cases we've examined so far, then, there are very significant financial benefits to government and industry from this use. In turn, theory then suggests that we would see significant influence on data, policy, and government decision making to maintain or enhance these benefits. So, we will take them in that order.

The Data: Driver History Record (DHR)

Governments are allowed – under the policies discussed in the next section - to sell various records falling under the category of "driver data," like vehicle registration information, all of which are of interest to the private sector. However, this analysis is limited to DHR's due to the large volume of revenue they produce and to help limit the universe of decision variables and related interdependencies by focusing on just one type of data. Data about drivers are collected by state Departments of Motor Vehicles (DMV) as part of the licensing process, and then on an ongoing basis in the form of a DHR. The driving history includes violations like running a stop sign or speeding, as well as accidents. The source of the data varies and while most originates in the court system, some entries are administrative and made by the driver control administration itself, like actions as a result of a DUI (driving while intoxicated) related to driving privileges (PRI-39). The record also contains general information that includes personal data, and restrictions that may apply (requiring eyeglasses to drive, for example). In today's world, the records exist in electronic form in a driver control system, with other data, such as violations, populated via an interface from the courts or other internal systems (PRI-40). The specific data and its retention varies across states (Sankey, 2015). There is also variance in how long violations remains on the record and the period of time covered by the records (Sankey, 2015). However, overall, because the critical entries on the record originate in law enforcement from traffic violations, exercising influence on the content of the records would be very difficult.

Policy on Access and Use of Driver Data

Today's market for driver history data has been significantly shaped by a federal law and its subsequent modifications, the Driver's Privacy Protection Act (DPPA), and by emerging substitutes for this data that are also available to industry. The DPPA, a federal statute as passed in 1994 and revised in 2000, restricts the distribution of "personal information," including "highly restricted personal information," associated with motor vehicle records held by state DMVs and their contractors (DPPA, 2014). These terms are defined explicitly in the Act (DPPA, 2014):

> (3) “Personal information” means information that identifies an individual, including an individual’s photograph, social security number, driver identification number, name, address (but not the 5-digit zip code), telephone number, and medical or disability information, but does not include information on vehicular accidents, driving violations, and driver’s status.
>
> (4) “Highly restricted personal information” means an individual’s photograph or image, social security number, medical or disability information;

The triggering event for this legislation is cited as a reaction to the murder of an actress in California in 1989 by a fan who obtained her address from the DMV (EPIC, 2019), along with similar cases of abuse of private information. While a review of its history and various lobbying efforts around it could be helpful in understanding the origin and evolving structure of the market for this data and the shape taken by the ecosystem around it, I have limited this analysis to exclude this simply for limitations on space – a chapter or more could easily be devoted to the DPPA. However, before going further it is important to emphasize that connecting personal information to risk is key to the commercial value of these records. The Act does not govern distribution of the data, but only the association of personal information with it when distributed. One can see, however, that because it is a driver that is insured, associating this personal information with the records is fundamental to its value to the insurance industry.

To address the privacy of personal information in driver records, the DPPA sets out a list of fourteen “permissible uses” under which personal information “may be disclosed” by state DMVs (DPPA, 2014) in association with driver records. It also contains provisions addressing resale or re-disclosure of information, requiring that requestors keep records of who received the information for five years and make them available to the state DMVs upon request (DPPA, 2014). The Act also includes penalties for non-compliance. The permissible uses can be grouped as follows:

- Related to law enforcement or other legal state uses, including driver safety and vehicle theft and litigation, and to provide notice of towed and impounded vehicles;

- By the motor vehicle industry such as for recalls or advisories;
- By businesses to verify or update personal information submitted by employees, agents, or contractors with their permission;
- By private toll companies in connection with their operation;
- For research activities with restriction on disclosure and contact of individuals;
- By licensed private investigators or security services for purposes within the fourteen conditions; and
- For "bulk distribution for surveys, marketing, or solicitations if the state has obtained express written consent" of the individuals involved.

There is also a blanket provision allowing disclosure where an individual grants permission. Finally, there remains one substantive condition that is most closely related to this research (DPPA, 2014):

> "(6) For use by any insurer or insurance support organization, or by a self-insured entity, or its agents, employees, or contractors, in connection with claims investigation activities, antifraud activities, rating or underwriting."

Even without an exhaustive look into the history of the Act, one can see that permission was carved out for certain industries and it is the insurance industry that, of course, has a significant stake in the driving behavior of its customers. This use is also one of two (the other being for insuring commercial drivers) out of four that, for highly restricted personal information, do not require the consent of the individual (DPPA, 2014) - with the other two being government use and litigation.

This federal law, then, shapes the market for driver history data in at least two ways: First, records containing this information may not be disclosed under any other conditions without the permission of the person to whom the record applies. In practice, this limits the private sector companies and business models that can participate. Second, the use of the term “may” in the law grants latitude to the individual states to further restrict access beyond these uses – and some states do. Along with these restrictions, they have also been allowed to adopt practices with regard to the downstream resale or redisclosure of information that are even more stringent, including preventing companies from recreating their own database of driver history of records. This might prevent companies from using non-current data, but also has the advantage of creating a dependence on the state to provide a recurring supply. A compendium of state law and policy on driver records (Sankey, 2015) documents the variations by states discussed here, a topic that is also addressed in the final section of this case.

In simple terms, this legislation creates a franchise in the form of access to information that other industries are excluded from obtaining for any of the commercial uses that fall into one of the categories listed (that the state adopts). While these laws and related policies might be difficult to change, it is possible that private sector influence could be brought to bear in this area.

State Decision Making and the Driver Data Market

The resulting ecosystem for DHR’s begins with the sources of violations feeding into a system at a state DMV, which may be located in a separate state agency, included as part of public safety or “highway patrol,” or transportation, or even a state department of taxation or its equivalent (Sankey, 2015). In turn, the resulting records are provided in electronic form to the insurance industry. In about half the states, this is accomplished through an intermediary that charges a fee as part of managing the distribution. There are also “products” created around this data, such as services that monitor for changes in the records (Driver, 2020). The companies buying these records may not be “household” insurance company names, but instead conglomerates, some international, that broker (resell and redistribute) information from many sources to many parties, including – in this case – insurance companies or other individuals or organizations who have “permissible uses” for the data.

As we move closer to ground level, however, the trail on which we seek evidence of dependence and evidence grows muddy with the complexity of differing schemes of discretion spread across different interests and incentives in state government. In examining the lay of the land in the 50 states through the lens of Sankey’s (2015) compendium, there are a number of variations involved:

- **Permissible Uses:** The degree to which all permissible uses outlined it the DPPA are adopted by the states differs, as do their rules for downstream reuse and re-disclosure.
- **Data Delivery Methods:** States vary in which services or data “products” they provide (individual; in bulk; as part of monitoring service).
- **Fees:** The fees they charge, and for what services, vary and these fees can change. The fees may be set in statute, or regulation, or simply be set by agency policy, each assumedly coupled with their own politics and barriers to – or ease of - change. A quick look at the wide variation – as of 2016, electronic copies ranged from $2.00 to $27.50 each per state (BRB, 2016) – seems indicative of variation in the related factors and decision making processes in the states.
- **Use of Resulting Revenue:** While not cataloged across the states, from the interviews I conducted (recounted below), the uses of the revenue from these fees also differs – some going to the agency that sets the fee, some fees included on behalf of other agencies for services unrelated to the programs of the agency, some directly to the general revenue fund

for the use of all agencies via the appropriation process, each with their own rules and restrictions (Sankey, 2015).

- **How Decisions are made (and who makes them):** As noted above, the DMV's reside in different agencies, the beneficiaries of the revenue may be in different agencies still, and variation in discretion over fee-setting means different parties, or groups of them, may be involved in these decisions.

The decision making environment is further complicated by the potential impacts of decisions about permissible uses, services and prices on the behavior and interests of the customers that make up the market for this data. As we will see in the interviews that follow, states must weigh the opportunity to maximize revenue and the need for program funding against anticipated risk of changes in price on the overall revenue produced from the market, given a particular service mix, and the ease of making these changes.

Challenges in Charting Influence in the State DHR Data Market

Given these variables and the number of states, adequately assessing the motivations and factors behind, say, variation in price or policy is complicated, and, from my interviews and attempts to talk with the private sector buyers of such data, I was left with a sense that they would like these practices to remain undiscussed, and that history and variation in them was not always well understood by those currently involved. The users of the data are not readily visible – that is, comprehensive annual financial reports do not include funds readily labeled as data sale (vs. fees), and the place where decisions about changing the fee are made is not easy to discern. What is known is that – as referenced in the introduction to this section – is that this is a business involving the receipt of hundreds of millions of dollars to states each year and that this fact is not always called out in state budget, revenue, and expenditure reporting in a clear and comparable way. An example from a conversation with a former administrative officer in charge of the budget of a state DMV, while only an anecdote, addresses the degree to which these types of use or allocation/reallocation aren't externally visible:

> That detail, by the time it gets to...overviews to the budget committee - anything that's in print - all you're going to be able to see is that expenditures from the General Fund decrease by a million and expenditures from [a fund into which fees related to electronic driver records are receipted] increase by a million. You're going to have to really dig in to detailed spreadsheets to get to see that property valuation had their state general revenue fund cut..." Honestly, in terms of any documentation, you'd almost have to compare internal agency budget workbooks from one year to the next. [For example] we may take a state general revenue fund cut in Tax and move money from [the fund] to cover it. You're not really going to see that at a detailed level. (PRI-08)

Industry challenges. After looking at the complexity of the ecosystem, however, one can easily take away a few things: 1) On the private side, companies are challenged by a myriad of rules concerning use of the data, with associated penalties for non-compliance at both the state and federal level; 2) As noted, fees charged for the data also vary considerably as does the apparent calculus of decision making about them - yet they are a significant cost of doing business for the companies; 3) To a large extent, because the state is the only source of the data, it is a "seller's market," which normally means demand would be relatively inflexible in the face of increased prices – absent ready substitutes- but also that this cost must be passed on to downstream customers who could have their own substitutes, ad infinitum. That is, the state participates in a market place made up of buyers, but the buyers may well participate in another market downstream where they are sellers.

Government challenges. State governments are faced with the challenge of – if this is a key revenue source to them for some purpose, which may not always be the case – maximizing revenue according to the variables identified earlier while maintaining good relations with businesses that provide that revenue, who, after all, are consumers of government services. The key fact underlying government dependence is that, in fact, insurance companies, for example, are under no obligation to buy driver history records, nor to buy them at a certain frequency or price. This uncertainty is likely to be a source of risk to government, and, to the degree the revenue is critical, fosters them to act to protect their position when threatened. However, to the degree market logics suggest that industry must use these records to maximize profit, they, in turn, are dependent as well.

Information Intermediaries. Enter into the complexity above another type of organization, the **information intermediary**, a company that sits between the state and private companies (as we know, some of whom themselves serve only as intermediaries to further distribute the data). While there are a few situations where other companies may serve this role, in over 20 states (NIC, 2018) a single publicly-traded company, NIC, has contracts to provide services to states that, among other sources, rely on a portion of the revenue these records generate to coordinate and deliver these service offerings, providing advice on fees, and generally negotiating this complex layer of laws and relationships that form the information market for state DHRs. While this company is not the primary subject of this research, the role of intermediaries should be considered in analyzing dependence among parties in this ecosystem. The most common situation, for example, set up contractually by the intermediary is that in return for managing technical, legal, and revenue collection responsibilities for this data for a state, the fee they receive also funds work by the company to provide other electronic services to the state – some for a fee and some subsidized in part or entirely by the revenue stream from the contractual fee from DHR sales. It is complicated, but well-summarized in their 2018 10-K Annual Report (NIC, 2018):

> Under the transaction-funded business model most commonly contemplated in these master contracts, our subsidiaries earn revenue through transaction fees paid by users in

> exchange for access to the services that we provide. These charges support the operation and maintenance of the services, as well as compensate our subsidiaries for the up-front investment and ongoing costs incurred in developing and maintaining the services, all costs that would otherwise be incurred by the state. *Our subsidiaries also utilize a portion of the revenue from these fees to develop additional digital government services that cannot be supported through transaction-based funding, either because the service would not have sufficient use, or the type of service is not compatible with charging a fee* [emphasis added].

This model, then, also has the possibility to create additional levels of government dependence to the degree that the DHR revenue subsidizes other government applications "that cannot be supported through transaction-based funding."

***Intermediary Financial Dependence on Driver History Record Revenue*.** Could a state sell this data directly without an intermediary? Some do. It is also clear that there is a dependence by the intermediary company on the state – and on the private sector - for this business. Here is an excerpt addressing this risk from the "Risk Factors" section of the same Annual Report (NIC, 2018):

> A significant portion of our revenues is derived from data resellers' use of our services to access motor vehicle driver history records for the automobile insurance industry. Transaction-based fees charged for access to motor vehicle driver history records in various states accounted for approximately 29% of our total consolidated revenues for the year ended December 31, 2018. One of these data resellers, LexisNexis Risk Solutions, accounted for approximately 19% of our total consolidated revenues during this period, or approximately three-quarters of our revenues from motor vehicle driver history records. In addition, approximately 15% of our consolidated accounts receivable were from LexisNexis Risk Solutions at December 31, 2018. While fees charged for access to motor vehicle driver history records are currently expected to continue to account for a significant portion of our consolidated revenues for the foreseeable future, regulatory changes or the development or increased use of alternative information sources, such as credit scoring, could materially reduce our revenues from this service. Our contracts with data resellers generally may be terminated at any time after a 30-day notice and may be terminated immediately at the option of any party in certain circumstances. Furthermore, our credit risk may increase in the event any data resellers experience liquidity or solvency issues. We generally do not require collateral to secure accounts receivable.

The role and dependencies related to the intermediary are important, but, as one can see simply from the earlier description of the decision-making environment, they sit within a complex ecosystem, where information about the revenues, operations, motivations, and relationships is not readily - if even publicly - available. In turn, as stated earlier, my experience in exploring this environment is that the subject is not widely visible, partial understanding abounds based on role and experience of those interviewed, and that parties who make use of the resulting revenue, especially those in the private sector, are not interested in providing detail about their incentives, strategies, or operations, all making the resulting picture produced by this research necessarily far from complete (see “Nothing to See Here: Changing the State DPPA” below for a description of a situation indicative of this challenge).

Participation in an Information Market and its Consequences: Impacts on Government Policy from Secondary Use of Driver History Data

Even with the caveats and complex factors related so far, of which there are many, there is still information to be gleaned about actual and potential impact of this mutual dependence. When government chooses to participate in an information market, the commodity in question is by definition a byproduct of a program, hence the term “secondary” use.” However, to the degree that it produces some benefit, including a financial one, to government, it is possible that not only the data-producing program, but others may come to depend on these benefits, even if their operations are at some distance from the origin and purpose of the data itself. In situations that follow, we see the complexity that can be introduced by the interaction of dependence and market incentives. Loss of revenue and corresponding price increases, substitution of other forms of data, and internal negotiations among government agencies – all of these can factor in to decisions about policy and practice.

Responding to Changes in Demand for Data: Managing pricing, dependence, and optics. The purchase of driver history data is optional, introducing risk into government dependence on this source of revenue. Depending upon the state, there may be multiple units of government that benefit from this revenue, as well as a private intermediary that manages the sales and distribution to the companies who buy it. And thus, when revenue falters, the state must act.

In one state, an employee at a private intermediary that had contracted with them recounted the following story. A fee from each DHR sold went to the DMV to support their new system, another fee to a state-sponsored intermediary (his company) who resold the records – more on that below - and then the majority of the revenue goes into the state general fund (PRI-09). He said that the state first began selling this data at the higher end of the pricing spectrum across the nation, but has maintained a fairly static charge for the records over time. The fee supporting the new state system was temporary to begin with, but as the project and costs have continued, it is "looking likely to be permanent." These fees are provided for in statute and thus required legislative action to modify (PRI-09).

In reviewing the evolution of the rate charged for the records, the employee highlighted an inflection point:

> In one year, the state saw a huge - the start of the significant decline in the sale of motor vehicle records. That decline became so rapid, and so devastating to our enterprise support that we went back to our oversight committee and requested that we raise the rate. (PRI-09)

Essentially, the losses to the operation of the intermediary selling the records on behalf of the state, coupled with the state's shortfall in their share of revenue from this activity, were unsustainable. Note that "enterprise support" is a euphemism for not just the going concern of the company's participation in the contract, but the other services across the "enterprise" of state government that the intermediary supports via the subsidization mechanism outlined above in the excerpt from a company annual report. Given the situation, the intermediary requested an increase in the price they charged per record to stay funding neutral, offering their oversight committee two choices, he said: “Look, we can put a fee on (vehicle) registration renewal, or I need a fee increase on MVR sales." Historically, there had <u>not</u> been what they termed a convenience fee for vehicle registration. And, of course, those impacted by a vehicle registration fee would be the public at-large, whereas the fee for purchasing copies of MVRs (DHRs) “falls on very few people” essentially insurance companies and the data broker industry (PRI-09). The committee granted the MVR increase, as, in the words of the employee “They did not want to have the optics of the citizen increase.”

*Substitute Sources of Data: Consumer Credit Reports***.** When asked about the sudden, precipitous decline in the demand for records, the employee addressed this frankly. “I don’t have any issue discussing it. We believe the decline started mainly because of changes to rules that allowed insurance companies to use credit scores as a form of rating a driver's liability” (PRI-09). He continued:

> I know that, to a consumer, a credit report is like a seven-dollar or eight-dollar item. I don't know what it is when you start buying them thousands at a time as I'm sure some of the large data miners do, but it's a lot less than the price we charge for DHRs. So, we believe the decline started for this purpose. A lot of the insurance companies - and we saw firm evidence of this - for existing business, rather than pulling the record every so many years when the renewal came up, they would rely more on a credit profile to determine whether or not they needed to change or reevaluate the risk they were undertaking in the rate. (PRI-09)

There are other impacts. In some states, there has been legislation to prohibit insurers from using credit reports in automobile insurance rate setting (Morton, 2016). Other arguments for this prohibition can be made, but is supporting this revenue stream part of the motivation for these laws?:

> Yeah, we've pushed such legislation, but it was not adopted simply because I believe, this is where I'm going to be very careful - I can't speak for members of our [legislative body]. From what I understand the general opinion is that insurance companies are their

> own thing….As long as the insurance companies can come in with their actuaries, and say, "Hey, look, here's a correlation - it's mathematical, it's there" - the states are going to be very hard pressed to come back and say, "You can't do that any longer." They would have to in some way prove that it was somehow biased towards a certain segment, which one could argue while the aging population, or the impoverished population are unfairly penalized by that. Unless they can find a fatal flaw in the system, the insurance company is going to be left to do what the insurance company is going to do. It's all about cost with them and risk. I know in conversations I've had with some of our data consumers certainly the driving record is the preferred method for rating the liability of the driver. But at a certain price point the risk of using another form of valuation simply looks better. (PRI-09)

Here we have seen yet another aspect of dependence, one that even results in promoting legislation to protect the associated revenue by preventing use of lower cost alternatives available in the information market. Yet, information markets are fluid and other substitutes exist - sometimes within government itself…

*Substitute Sources of Data: Court Records***.** The mechanics of private information markets have other ways they can seep into the practices of the state. The situation grows more complicated, for example, when an additional private intermediary - one not contracted to government - enters the market – in this case re-selling a substitute record and also seeking to optimize revenue for itself <u>and a different government organization</u>. We then see this behavior countered by efforts to coordinate distribution and pricing to maximize overall revenue.

To explain further: Ultimately, most violations on an individual's driving record begin as a violation recorded by the courts. The employee at the intermediary with an existing contractual agreement continued his discussion of the threat of substitutes to revenue from DHR's:

> We also know that our data purchasers are purchasing judicial information. And this also is something that I just read a recent article about - it's becoming a lot more common in that judicial information is usually less expensive than the motor vehicle records. They contain a lot of the same data, because what insurance companies are mostly interested in as well as all the other, not just insurance companies, they're all interested in violations, suspension, accidents, et cetera. They can then purchase that information from the judicial system which is far cheaper. What's interesting in the article that I read is that why that has not been cost effective in the past is because in most states the courts are

> dispersed either on a county or municipal level. Rather than dealing with one state entity they need to deal with lots and lots of jurisdiction, and there's actually been some companies that have sprung up, they've taken on the burden of saying, "Well, we'll work with all these individual judiciaries, we'll aggregate the data, then we'll sell the data back." Court data is also being used for monitoring records – and we know that as a fact." (PRI-09)

There is interest on the part of both the intermediary and the executive branch in addressing this "problem," framed as potential revenue available in the market that is being given up by lack of coordination by the state in the supply and pricing of its records. The intermediary employee continues:

> In our state, the judiciary is very central and so the purchase of that data has definitely continued to hurt our sale of driving records. We're working with the administration, because - while we talk about funding for our operation (the intermediary) - the state has lost millions of dollars, or potentially millions of dollars, over the past few years (from the price being undercut). Which from our state's budget standpoint is a significant amount of money.
>
> The issue with the branches of government, the separation of power is very, very strong and politically fractured in our state. The executive branch can't go there and say "Don't do this anymore." They can ask, they can paint a picture, and we're working to do that, the one big tent which is what the public really wants. But that's an ongoing process. (PRI-09)

The Perpetual Proviso: Additional Considerations and Impacts due to Dependence on Revenue from Driver History Data. While the examples above show the state and its intermediaries seeking ways to recoup or increase revenue, there are other dimensions to the equation. Like the state discussed in the previous example, another state instituted a similar fee to support the ongoing cost of the system used to house the driver data. While the fee was delegated to executive management of the agency to set (and change) – rather than the state legislature - there was statutory language that ensured it would be used only for this purpose – to start. In speaking again with the former administrative officer referenced above, who had spoken about the opacity of internal budget impacts within an agency, the story grows more interesting. As he noted:

> The sale of these records in bulk and individually constituted a fairly significant part of our operating revenues. So, on the budget side, I certainly was interested in those rates and when they changed. There are two circumstances (where they might seek a fee increase). Let's say legislation is being considered that would cause a new record type (to be created), so we have computer costs and ongoing support costs, operational costs - whatever the fiscal impact memo (Note: Budgetary estimate to the Legislature of a proposal's impact) may be, we would typically point toward the operating fund that supports that operation. So, yeah, if it's going to affect that, then we might ask for more - we might propose to raise the fee to cover those costs.
>
> So, legislation / new responsibilities could be (the cause) - but it is the least likely scenario. *The primary one - the only one I'm really familiar with - is budget cuts in other areas* [emphasis added]….For example, we have state general fund – essentially income tax revenues -cut out of [a division]. We compensated for that by increasing record fees and bring in more money and balance that loss, that cut we had. (PRI-008)

Here we see further evidence, similar to the revenue crisis referred to in the previous state, that revenue from the sale of government information can be integral to an agency's fiscal operations, in this case resulting in fee increases, and also that agency programs unrelated to the production of the data can depend on this revenue. While the details of this dependence may not always be available to the larger policy-making environment, on a practical level, it may sometimes be visible to legislators and state budget officials. In this state, this is illustrated by a willingness by legislators to "temporarily" suspend – each year - a long-standing statute that specifically limits the use revenue from data sales to support only expenditures for the technology needs of the organization to allow the revenue to be used for broader purposes:

> There have been at least two occurrences where it was actually part of the Governor's budget recommendation for the department to increase record fees to offset general fund cuts - so it is pretty transparent there. Beginning in the late 90's, early 2000's, for example, the trend began to directly use those monies to offset general operating fund support. So, you're looking at a 17-18-year history of using provisos (a temporary legislative action lasting only for one year) to override the base statute for (these funds) and say they can be used for general operations of the department. You've sat through some of those arguments about how long can we use a proviso? Well, these have been in

> place for at least 18 years. The proviso submitted in the budget every year has said - not withstanding those restrictions - the department can use this money for anything. (PRI-08)

On an operational level as well, agencies understand they are participating in an information market. For example, their decisions are sensitive to the potentially negative impact of raising prices on demand – and thus overall revenue - especially given the availability of substitute goods (credit reports or court records) we've seen in previous examples. And continued good relations with their "trusted business partners" (the buyers) matter, too. He continued:

> You have the two groups. The bulk buyers obviously never voluntarily say "Oh, please raise my fee." The individual buyers which are small volume buyers, which are the bulk of our money, obviously don't have that kind of "800-pound gorilla" force behind them. We try - probably the best benchmark to make sure we're not going to actually see a reduction in record sales because our price is too high - to stay consistent with what other states are charging. So, that it's not - you know - yes, people will squawk, but this last time it was raised, I think we were projecting and hoping to see about a million dollars in additional operating money, and for whatever overarching reasons, we saw double that - sales did not decline.
>
> You can't change the fees completely in the dark. And, obviously, if we're going to raise the fee, we've done whatever - research with other states, with major vendors. One lesson over my twenty years is "Don't surprise people with a rate increase and then wish you hadn't imposed it." Certainly, we've never let it get to the point that I'm aware of that when we were running our own regulation for a rate increase or whatever that they came to legislative committee, for example, and lobbied in force against it. We work those things out. Generally, they have to be trusted business partners just as much as they're also vendors. (PRI-08)

This state, too, had to deal with internal tensions related to the resulting revenue from secondary use, in this case from situations where other state organizations wanted part of the revenue:

> The bigger discussions tended to be when there were external entities trying to either take a piece of our pie, or increase the total fee to some degree that we felt might threaten our revenue stream. Then we would become actively involved in those discussions helping other agencies understand why you might not want to try to suck any more money out of this or that industry. (PRI-08)

The dependence across programs can manifest itself in other ways, too, based on the funding and sensitivity an agency or its program may have to the information market (and associated revenue) risk. In a third state, a person working for their intermediary explains:

> [Our state] is a little unique. I don't know that you'll find this in a lot of states. The majority of the statutory fee goes to fund law enforcement retirement. The [related department] has a very vested interest in DHR revenue doing really well, as it's supporting law enforcement retirement…The state is so dedicated and vigilant on where the revenue goes against the law enforcement retirement. They don't want to do anything that will affect that revenue stream. (PRI-10)

One way to potentially expand revenue is by expanding the types of service that a customer can purchase. Instead of an insurance company buying the record about a customer they insure every quarter, in some states, an additional fee-based service is offered that will monitor a list of their customers' records for changes. It costs less, and avoids the requirement to buy the record, only to find out there have been no changes / violations in the previous period (see reference at Maryland.gov, 2020 above). There are variations – sometimes a customer is required to buy the record when a change is detected (Sankey, 2015). Regardless, the state or its intermediary adding this fee-based service can increase overall revenue by bringing in new customers. In turn, it introduces the risk that existing customers may switch to the potentially less-costly monitoring service. While still having to pay something for the new service, they may end up buying less records as a result, lowering the combined revenue produced across both services. Taking on that potential market risk was a no-go in this state and when asked about the source of this restriction and the reasoning, the manager of the intermediary responded:

> It's from the state. We have had several companies ask us about monitoring. It is from the state. We can't offer monitoring. We have the ability to develop the system for them, but I think if they did put in some sort of monitoring, they would be very strict with it and we'd get a quarterly review process of revenue. If we saw revenue decrease, then monitoring comes down...They just don't want to put anything out there that would affect that retirement fund. (PRI-10)

Considering Variation in Dependence and Influence on State Driver History Records

These brief stories of DHR sales in state government reveal several new aspects of dependence on government information by both the private and public sector. Unlike our previous cases, this information is sold and produces revenue directly for state governments. As a result, the dependence is perhaps more obvious – one would assume the private sector would not pay for data it didn't need. We see that law plays a role in shaping access by narrowing it, but also may be used to prop up revenue by restricting commercial alternatives. And, that it can be worked around to make sure the revenue gets to where it is needed. Decision-making in agencies – about fees, about delivery methods and services – may be influenced by supply and demand in the information market, as well as perceived risk (including "optics" to both the public and business) and reward. Finally, intermediaries may come into play as brokers and advisors in the design of rates and delivery, and – in the case of at least one prominent model – their role as a mechanism for cross-service and cross-agency subsidies may strengthen both their influence and that of the beneficiaries of the subsidies in ensuring the revenue produced from DHRs maintains their funding.

This high-level review leaves much more opportunity for exploration, especially – for comparison - with those states that charge the lowest rates and do not make use of an intermediary to sell data. In addition, those states that have additional restrictions on the purposes for which data may be used could shed light on how the balance between privacy and additional revenue is negotiated. The process of deciding upon or negotiating a rate change could be observed close-up, with interviews of the participants to better understand the range of considerations that come into play. However, while I was able to interview a small set of intermediaries in other states, and a few state people, the vendors purchasing these records were tight-lipped and have provided none of the information included in this study. A clue to their motivations and methods, if any, of influence is apparent with the switch to substitute court records or credit reports for DHRs that was discussed above, and there is no doubt much rich material awaiting research into the tertiary markets into which they sell the data and its uses, along with the motivations of those they depend on for such sales, the downstream users.

Nothing to See Here: Changing the State DPPA

In closing, perhaps further reinforcing my assertion of the opacity and lack of familiarity with the details surrounding this subject by some of the principals, I include the following story, where the motivations of the parties, their incentives, and the eventual impact of the resulting decisions are all a little hard to judge – yet the outcome was clear. It involves a vendor proposing changes to the law governing access to driver records, a policy change involving democratic oversight in the form of a legislative hearing. In considering my thesis that the influence produced by secondary use could serve to erode democratic processes, the events described here should be taken as, if not supporting evidence, an incentive to examine this idea further.

In an earlier section of this case, I discussed the federal Driver's Privacy Protection Act, noting that it includes language allowing states to adopt more – but not less – restrictive provisions for the use of driver records than the federal statute. This story begins in early February, when a data broker serving various markets that include the insurance industry, IHS Markit, sponsored legislation in one state to "sync up" a more-restrictive state statute with the federal DPPA statutes. One might wonder about the content of the legislative hearings and debate on the bill, given that the existing state DPPA did not include some uses that were allowed under the federal statute, especially given that access policy is an area where we expect to find evidence of influence exercised in favor of one or both parties. And - it seems - so we do.

House Bill 2179, titled *Adopting the Driver's Privacy Protection Act* – was introduced on February 7, 2019 (H. 2179a, 2019). Despite the title, however, a Driver's Privacy Protection Act already existed in this state's law. The bill struck out the provisions of the then current state law (K.S.A. 74-2012), leaving only four specific enumerations that were already permitted under different laws for assisting other units of government (Selective Service, Commission on Veteran's Affairs), assisting employers in monitoring employees that were required to drive in the course of business, and one that described a certain use of the records that was allowable, but was specified in greater detail, which also happened to cover the main use IHS Markit (formerly the directory company R.L. Polk that also owns Carfax) made of driver records, which was to use information derived from the records, absent names and addresses of current or previous owners, to help prepare vehicle history, the primary Carfax product (IHS Markit, 2020). It then added a provision that stated "(E) any other purpose authorized by the driver's private protection act, 18 U.S.C. 2721, as it existed on January 1, 2018." As a practical matter, these changes appeared to serve the purpose of bringing the legislation into alignment with the Federal DPPA.

At the outset of the initial hearing, two pieces of information were presented to the committee. The first, was an explanation of the bill by the staff of legislative research, briefly covering the background of the Federal DPPA, indicating that "the Act currently applies to all state departments of motor vehicles and all officers and employees and contractors of those departments." The analyst then read the federal permissible uses aloud to the committee. He continued "I want to make clear that some of these permissible uses are already in the statute we are amending, K.S.A. 74-2012, some of them are not. However, this federal act is applied to the division of vehicles right now, so really, these uses are already permissible, whether they are listed in 74-2012 or not." (Adopting the Driver's Privacy Protection Act, 2019). And, then, “Really what the bill is doing is deleting any reference to a permissible use that’s in the federal Act and then specifying in subsection c(1)(E) that it is for any other purpose by the Driver’s Privacy Protection Act as it existed on January 2018 – that’s the federal law.”

After the explanation was complete, a legislator asked if this was more or less a technical fix. The analyst said “You could say that, yeah. Like I said, all these permissible uses, the federal Act already applies to the department…this is just kind of lining them up.” Another legislator had a question “Is there a substantial change in any aspect of citizen’s privacy that’s being opened up here. It seems like it says that any state agency now has access pursuant to their jobs, whereas that might not have been there before. Am I reading that right?” The analyst explained that it was there before and called his attention to where it was in the existing legislation. “The insurance piece, that’s not new, either?” he continued. The analyst said it was not. And, a third legislator was recognized to speak: “So, why was this introduced? Is there something that I’m missing – and why these corrections? Was it misused or something? The analyst said he would leave that to another conferee to answer, as far as the intent (Adopting the Driver’s Privacy Protection Act, 2019). There were no further questions.

The second piece of information was written testimony from IHS Markit (IHS Markit Testimony, 2019) that was presented by an associate director at the company. He introduced himself and explained the connection to Carfax, the name by which most people would know them. He explained that the company was here in support of the legislation and that the objective was “to bring the Kansas state statutes consistent with the federal DPPA, to clean up the language to make the language consistent. There are more than 40 states who adopted the federal DPPA as it existed, and as it continues to exist, and in order to make this clear, we are encouraging this public policy to eliminate any confusion, any unnecessary litigation, or anything that anyone who uses the information may have." He continued "So, we have been in business for 140 years, we have been buying state information for almost a hundred years, and we have a contract with the state of Kansas to acquire the information. We just want to make this technical change, this technical fix, to the state statutes" (Adopting the Driver's Privacy Protection Act, 2019). He went on to further explain the value of Carfax services. The written testimony submitted for the record was about a page in length and said essentially the same thing, including the statement that "This technical amendment would ensure the continuity between the Federal DPPA and the Kansas state statutes, and the majority of other states" (Adopting the Driver's Privacy Protection Act, 2019). The final question from a legislator was whether or not it would allow robocalls and he strongly confirmed that it would not. The chair ended the discussion by saying "You got off well" to the IHS Markit executive's laughter, who responded "I hope so." There was no neutral testimony nor opponents on the bill and the hearing was closed, the bill then recommended for passage by the committee, placed on the consent calendar and approved 124-0 by the House (H. 2179b, 2019).

The bill did not receive a hearing in the state Senate. Instead, language of a bill (HB 2126) regulating electric scooters that had already passed both houses was replaced with the contents of HB 2179 by a joint conference committee, whose report recommended passage. Both houses voted unanimously to accept the conference committee report and the bill was subsequently signed into law, where it remains in effect today (H. 2126, 2019).

Inaccurate Testimony? States May Adopt More Restrictive Provisions. From testimony and discussion in the hearings, the rationale for ensuring "the continuity between the Federal DPPA and the Kansas state statutes, and the majority of other states" is never quite clear. Industry might be served by continuity, one supposes, as they testified, just to clear up any misunderstanding if there were ever any questions. Yet, it seems odd – why now, given that no barrier to current operations is ever presented. And why the language about the "majority of other states"? While industry's motivation and that of the state agency whose data sales would be affected may be unclear, there appears to be a "misunderstanding." As summed up in the 2015/2016 MVR Access and Decoder Digest (Sankey, 2015), a guide for professionals in the driver record industry:

> While the DPPA implemented a minimum set of standards, states' privacy rules can be and often are more restrictive**.** *States may choose not to adopt all 14 "permissible uses" listed in the DPPA, or may choose to adopt even more stringent policies* [emphasis added]. Each state chapter within this book indicates which of the 14 permissible uses that a particular state has adopted or NOT adopted, and if the state has stricter rulemaking.

Similar references are made in other publications and websites about the use of driver data, including at the Electronic Privacy Information Center (EPIC) under the heading "State Protections May Be Broader than the DPPA" (EPIC, 2019). As evidence, one may look at the current statute - Alaska Administrative Code Section 28.10.505 (AAC, 2020) - for the State of Alaska to see that it does not include use by private investigate agencies, for example, or the Arkansas Code Annotated Section § 27-50-906 (ACA, 2020) to see that the statute is missing several of the permissible uses included in the Federal DPPA, including those mentioned above that would be, in effect, added to state law by the proposed legislation. There are at least several permissible uses that had not been available in the state under the old law, including licensed private investigative agencies and licensed security services for any purpose permitted by the DPPA, for use in connection with private toll transportation facilities, notice to owners of towed or impounded vehicles, and for research purposes as long as personal information is not published or used to contact individuals (DPPA, 2014).

What does all this mean? It is hard to say, as there is no straightforward way to find out a) if new parties are going to use the data, or b) if they may have been discovered to be intentionally or inadvertently using it all along, with the need to get the law "aligned," or c) if there is yet some other impact, or it is simply as it was stated. Nor is it clear why the attorney for the department of motor vehicles would not dispute the legislative analyst's or vendors' explanation of the federal DPPA's applicability, nor why the legislative analyst apparently either took the explanation by the vendor and/or the agency and never looked further into it. Once can imagine the discomfort in asking them those questions, yet, the documentary record remains for future research – in the archive.

Conclusion

In the three cases in this chapter, we watched as the benefits to industry from their secondary use of government data increased, as did the benefits to government – and the breadth of their impacts. First, a company, for public relations value, some apparent altruism, and perhaps an increase in revenue from that goodwill and publicity, tried to encourage government to "make things easier on themselves" and accept the company's offer to help them achieve their goal of better public health outcomes by simply making their data easier for the company to use. Yet, several forces, from the influence wielded by the restaurants they regulate, to the visibility it would bring to their inspection practices, and the lack of any benefit to them for standardizing data for a national purpose that pertained mostly to people who lived there – and, frankly, the poor understanding the company had of even the existence of these barriers and how to address them – mean that there was little interest, and thus dependence, by government on the company's use. We will see other cases where government has limited benefits, but the barriers identified here are important to carry along was we build our understanding of the phenomena of dependence.

In the second case, we learned two things. One, that faced with a lack of funds to accomplish core duties, an agency may seek and find alternative ways to accomplish them, and be willing to justify less than perfect outcomes in the name of progress. In turn, we see our first instance where market values show their ability to be transmitted via dependence. The resources provided by Ancestry don't just come with requirements for a period for exclusive rights to sell data, they're focused on helping government perform its function only for specific types of data that can produce revenue for them – another lesson to keep in mind as we progress. And, I leave that case posing a question as to how far this combination of private sector interests and government lack of funding might more negatively affect their mission - and history itself – by leading to cases where the data to be retained, not just digitized, begins to take into account its market value for survival.

The final case addresses one of the largest and most lucrative markets known for government data, and, for our purposes, it shows. Governments in this case are now clearly participating in an information market, one so rich that a company (the private intermediary) is even attracted into a market where they can make a profit just from helping government make money! Decisions about pricing must be made, and risks to the funding of multiple, even unrelated, programs must be weighed in the process. Interests and competition for this revenue inside government and even the private sector that government regulates must be coordinated, or even stymied, to ensure it keeps on flowing – all the while keeping good relations with the private companies who provide it. This world is much more complex than the two we saw before it, so much so that even legislative staff who analyze the law around it make mistakes, and changes proposed by industry that open up access to private companies for more uses are agreed to with little fanfare, or even notice.

In these three cases, then, we begin to see that there is something here in the idea of mutual dependence on the secondary use of government information. In the next chapter, we will look at another giant market, this time at the federal level, where the data is open and no revenue is allowed – and the vendors are quite demanding. Like these cases, we find new aspects of dependence, and begin to see government take a more entrepreneurial approach to seeking benefits, if they can find them.

Chapter 3: Emerging Commercial Influence on NOAA Policy and Practice in the Secondary Use of Public Environmental Information

Introduction

This chapter examines commercial sector influence on NOAA information policy and business practices by tracing the development of its current public-private partnership policy and select activities of a committee created to operationalize it. In turn, it focuses on an entrepreneurial initiative inside NOAA that showcases the value of its data to industry and promotes greater use. The final sections map emerging relationships between those engagement efforts and two other NOAA organizational functions, one supporting the process of managing and prioritizing NOAA data requirements, the other charged with assessing and promoting the value of NOAA data. By looking at how commercial interests are, or are coming to be, reflected in each of these venues, we continue the study of information dependence by highlighting different paths, mechanisms, and incentives by which commercial interests come to influence the content, structure, and availability of NOAA data.

From the perspective of our research design, these two cases represent different extremes. In the first, industry has the upper hand, gaining concessions from NOAA that allow them to define what competition is – and stop it when it suits them. And, they are able to get them to adopt practices that ensure a reliable supply of data and that NOAA will consult them whenever it considers taking action to disrupt or change practices around anything they rely, or might come to rely on. After industry takes advantage of these policy concessions in a classic "easier to ask forgiveness than permission" move to run a high-speed pipeline into NOAA's data center to gain new access, they are greeted not with either of those responses, but instead cut off before they get very far by a legal threat – from yet another member of industry! From this we learn an obvious lesson: businesses are in this for their own – not each other's – benefit.

In the second case, the situation is much different. Here NOAA starts an effort with a contracted intermediary to visibly promote the use of their data, hiring a contractor to create reports and materials about the wide uses to which it is put in the commercial sector and, by extension, the value of the programs that create it – all in the service of increasing use, to the point of soliciting requirements from these users for how they might improve these "products." In this case, government seems to be cultivating influences and making changes on its own in hope of gaining more use, and sponsorship. But, there is another reason. They are in a unique position of holding critical information that could be used by companies to adapt to coming effects of climate change. Yet, the opportunity and resulting market is nascent. Government can only do so much in this area, and only has so much of a mandate for its own action. But, NOAA has come to believe, if they can use their vast data assets to help model the opportunities for the private sector, perhaps they will fill in the gaps and the market, solutions, and the hoped for resilience will come into being. In this case, then, NOAA is the protagonist, coming off the ropes in the previous case to look for ways they can serve their own interests – and the greater good, of course.

Defining Fair Weather: Embedding Commercial Interests in NOAA Information Policy

Our analysis of commercial influence on NOAA data begins with a report issued by the National Research Council in 2003. Titled "*Fair Weather: Effective Partnerships in Weather and Climate Services*," (NRC, 2003) the report was developed at the behest of the National Weather Service (NWS), an organizational unit of NOAA. It represents another step in a long line of reports and policies stretching back to the late 1940's that constrain the ability of the NWS to offer products or services that compete with the private sector while, in turn, also ensuring public access to the information the government produces so that it might be commercialized (ACWS, 1953; NRC, 2003). As of 2003, the most current policy was the 1991 "Policy Statement on the Weather Service / Private Sector Roles" (NOAA, 1991) that enumerated specific activities appropriate for each sector. However, the *Fair Weather* report was commissioned to, among other things "Identify barriers to effective interaction between these sectors" (NRC, 2003) and set in motion an effort to redefine this policy approach with two related findings. The first was the report's primary conclusion that "It is counterproductive and diversionary to established detailed and rigid boundaries for each sector outlining who can do what and with which tools" (NRC, 2003). The second is represented by the first recommendation in the report, which was to replace the 1991 policy with one that ***"**defines processes for making decisions* [emphasis added] on products, technologies, and services *rather than rigidly defining the roles* [emphasis added] of the NWS and the private sector" (NRC, 2003).

After the delivery of the report, a committee was formed to address this recommendation. The resulting policy, *NOAA Administrative Order (NAO) 216-112 Policy on Partnerships in the Provision of Environmental Information* (NOAA, 2006) is currently the central NOAA information policy addressing *provision* of data to external parties within the overall federal information policy framework. The following discussion identifies key themes in that policy that demonstrate the embedding of definitions and mechanisms that enable influence by commercial interests.

Boundaries of Convenience: Merging Government and Commercial Interests in Information Delivery

The document containing the partnership policy introduces the term "Environmental Information Services," language that expands its scope from weather and climate data, to water and "chemical, biological, and ecological parameters" (NOAA, 2006). Next, pursuit of the creation and delivery of these services accrues to an "Environmental Information Enterprise…composed of government, private sector, and academic/research institutions" (NOAA, 2006). While it is understandable that multiple actors both public and private may share common goals that benefit society, this terminology and the associated conceptualization can be used to selectively blur institutional boundaries that exist for reasons of democratic representation and control, perhaps with different purposes (environmental regulation vs. commercial enterprise) and different priorities (basic research vs. "use-inspired" research based on "practical" issues, perhaps of a commercial nature, as we see in the second case in this chapter). It is worth considering what benefits might accrue - and to whom – from looking at the relationship between government, industry, and academia in this way, along with possible effects on priority setting and decisions about which sector makes what investment. Redefining agency in the achievement of goals to be coordinated in the services of a homogenous "enterprise," while allowing a flexible, subjective view vigorously enforced by commercial interests of "competition" creates what I call "boundaries of convenience" that are an important aspect of facilitating commercial influence. That there should be no doubt that influencing resource allocation is a goal here, note that the *Fair Weather* committee charter requires it to make recommendations as to how each of the sectors can coordinate to "[M]ake the most cost-effective investments in needed infrastructure (and) efficiently share the information generated from that infrastructure" (NRC, 2003).

No Surprises: Ensuring the Reliability of the Information Product Supply Chain

The introduction to the 2006 policy reads "NOAA will not haphazardly institute change in existing information dissemination activities, or introduce new services, without first carefully considering the full range of views and capabilities of all parties as well as the public's interest in the environmental information enterprise" (NOAA, 2006). We find the apparent intent of this statement, however, in language included in the original draft of the policy that was later revised (NOAA, 2004a):

> No surprises: Unless public safety or national security concerns dictate otherwise, NWS will provide all users, including those in the private and academic sectors, adequate notice and opportunity for input into decisions regarding the development and dissemination of significant products and services, and their discontinuance.

We have in this statement clear evidence of a) government's recognition of industry dependence on a resource they provide, b) an agreement to create a mechanism that allows industry to depend on a reliable supply of information critical to their operation with minimal interruption, and c) ensuring commercial interests a seat at the table in decisions made both about new and existing sources of information supply. Remember that this is data produced by government as part of daily operations to perform its mission, with secondary "reuse" just that, and that it is provided free of charge to external users. If such a concession were made by one party to an agreement in a commercial setting, its contractual value as "consideration" would be immediately evident. For a business to be able to ensure a consistent and reliable supply of an information product or service core to its operations enhances its attractiveness and financial value to the market and reduces risk that no longer needs to be "priced in," increasing its viability and competitiveness. The commercial necessity of such guarantees in an information market is also re-emphasized in comments made by industry executives involved in the NOAA Big Data Project in Chapter 4.

Ensuring Access and Formatting for Additional Processing

The partnership policy commits to "open and unrestricted access" to data in "forms accessible to the public as well as underlying data in forms convenient to additional processing";

"in vendor neutral form" to "advance the environmental information enterprise," delivered in ways that "comply with "recognized standards and formats…to ensure data can be integrated" (NOAA, 2006). To briefly unpack this idea, NOAA – in informing the general public and ensuring its safety via the NWS, for example – seems unlikely to find it necessary to allow them to download or livestream data files for further processing. The former use is addressed by the "forms accessible to the public," the latter – and there can be significant expense associated with making data available in this way – is most certainly for the benefit of commercial (and some academic) users and demonstrates the extension of the commercial information supply chain requirements into government.

Cooperation not Competition: Preservation of a Secondary Market in Environmental Information

While the final version of NAO 216-112 was published in early December 2004 (after a six-month period in which they received 1,473 comments) (NOAA, 2006), a notice was published in the Federal Register just nine months later seeking "clarification to address apparent misunderstanding regarding the intent of the policy with respect to the role played by the private sector in the environmental information enterprise as a whole" (NOAA, 2004b). It continued: "The present policy does not adequately express NOAA's views of the critical role played by the private sector in the environmental information enterprise as a whole. NOAA is sensitive to the concerns and prerogatives of the private sector and has no intent to displace it" (NOAA, 2004b). The commercial sector's view of the gravity of the issues with the partnership policy was summed up in the following statement:

> Regrettably, the parent agency of the NWS, NOAA, repealed the 1991 noncompetition and non-duplication policy in December 2004. Its new policy only promises to ''give due consideration'' to the abilities of private sector entities. The new policy appears to signal the intention of NOAA and the NWS to expand their activities into areas that are already

> well served by the commercial weather industry…NOAA's action threatens the continued success of the commercial weather industry. It is not an easy prospect for a business to attract advertisers, subscribers, or investors when the government is providing similar products and services for free. (151 Cong. Rec. 6568, 2005)

An excerpt from a company letter to NOAA or their corporate spokesperson? Opening a new front in the exercise of influence on the partnership policy, these are the words of Senator Rick Santorum (R-Pennsylvania) in the Congressional Record from April 14, 2005, directly addressing Section 4 of the partnership policy while introducing Senate Bill 786 titled "The National Weather Services Duties Act of 2005" (S. 786, 2005). The proposed legislation would prevent NOAA from providing "[A]ny product or service except those forecasts and warnings designed for the protection of life and property by the general public…that is or could be provided by the private sector"; insured that data would be available "real time" to external parties "without delay for internal use"; with "all members of the public hav(ing) the opportunity for simultaneous and equal access"; and that data be provided through "data portals designed for volume access by commercial providers" (S. 786, 2005). Perhaps a coincidence, but Santorum's state of Pennsylvania includes the headquarters of Accuweather, a large weather services company (https://www.accuweather.com/en/about/). The Santorum-sponsored legislation was not the only political / lobbying effort in evidence. In just one example, the National Council of Industrial Meteorologists (NCIM) submitted a comment (#91) (NOAA, 2005) on the proposed clarification stating that they had developed and circulated a position paper on the need to "fairly and forcefully" reduce competition with the private sector and had also met more than once privately with NOAA officials, including a meeting in Washington, DC shortly before the request for clarification was issued, "clearly reflecting NOAA's willingness to listen and respond to its partner constituency." Lobbying then, both with the legislative branch and NOAA executives, is a mechanism of influence used by industry to impact government policy on information use, access, and distribution. The result? The current (NOAA, 2006) policy reads "NOAA recognizes that *cooperation, not competition with the private sector* and academic and research entities best serves the public interest…" [emphasis added].

Opening Up Decision-Making

In line with its prioritization and decision-making being subsumed into a negotiation with the larger enterprise, NOAA commits in the policy to "open consultation" via "orderly processes for seeking input and suggestions to create, modify, or discontinue products and services" (NOAA, 2006). They also commit to seeking advice "in accord with the Federal Advisory Committee Act (FACA)" (Public Law 92-463). While NOAA has several advisory committees that fall under the FACA, in the next section of this case, the practical implications of this statement with regard to the NOAA / commercial partnership will become clear when a working group of the Science Advisory Board (SAB), a FACA committee, proposed aggressive steps toward direct access to NOAA computing resources and collaborative participation in NOAA system design, both under the idea of open consultation and access.

Equal, but Different: The Application of Equal Treatment by Class and Competitive Advantage

While each of the provisions discussed above embed commercial interests in the policy, another provides a telling framing of the issues latent in the concept of equality in commercial access:

> Equity: NOAA will be equitable in dealings with various classes of entities and will not show favoritism toward any particular entity within a class. NOAA recognizes it has special responsibilities to some users (e.g., public safety officials) and different legal requirements for its interactions with entities of different types (e.g., other federal agencies). NOAA will not provide an information service to one entity unless it can also be provided to other similar entities. (NOAA, 2006)

While this definition promises a level playing field for access to data among “classes” of commercial competitors, there is not a specific precedent for this approach in the overarching federal information policy, OMB A-130 *Managing Information as a Strategic Resource* (OMB, 2016). However, one of the principles included in that policy is that agencies should “(5)(e)(7)(d) Consider(ing) target audiences of Federal information when determining format, frequency of update, and other information management decisions;” (OMB, 2016). That said, this phrasing reveals a practical distinction in the concept of open access to information introduced by the availability of machine-readable data. While government policy emphasizes that data should be equally accessible to all, in practice, significant financial and computing resources, along with technical skills are required to receive and make use of large amounts of machine-readable data, especially in “real-time.” This (re)definition of equity does two things: First, it advantages commercial users for whom more sophisticated methods of access can be made available by freeing the definition of equality from being anchored to the lowest common denominator of all users. But, second, as commercial firms vary in their technical capabilities and resources, it opens the door for conflict over unequal access among commercial organizations who depend on secondary access to NOAA data. This approach to equity and its results will continue to echo throughout this study.

From Policy into Practice: The Science Advisory Board’s Enterprise Information Services Working Group (2009)

NOAA Science Advisory Board’s (SAB) Environmental Information Services Working Group (EISWG) was first convened in mid-November 2009 (NOAA SAB, 2009). The organization is a standing working group of the NOAA Science Advisory Board, a Federal Advisory Committee governed by the Federal Advisory Committee Act (Urhart, 2009). Their distinction as a "working group" versus an advisory committee subject to FACA is important, as it is not subject to the same level of administrative reporting, assessment of conflicts of interest, etc. under the FACA (Public Law 92-463). The EISWG at this time was made up of weather industry representatives (like Accuweather and Weather Services International) or technology firms with industrial meteorology interests like Harris Corporation and Raytheon, along with academic and government members (NOAA SAB, 2010). The work products of the EISWG are presented to the SAB, who then reviews and formally issues them.

The EISWG was established by the SAB at the suggestion of one of its "ad hoc" committees (NOAA Partnerships Group, 2008) convened after the implementation of the NAO 216-112 partnership policy discussed earlier, and was based on a recommendation in the *Fair Weather* report (NRC, 2003). Its initial "terms of reference" were to:

> 1) Provide advice on improving communication among the sectors (government, private, academic), 2) Provide advice on incorporating scientific and technical capabilities to enhance NOAA products and services, 3) Provide a sounding board regarding implementation of NOAA's Policy on Partnerships in the Provision of Environmental Information, and 4) Evaluate NOAA effectiveness in responding to advice received from the EISWG, and the environmental information enterprise as a whole. (NOAA SAB, 2009b)

The EISWG wasted little time in moving forward on its second charge of providing advice on enhancement of NOAA's products and services.

Moving the Boundary: Toward Open Environmental Information Services

"Therefore, the basic position advocated is that NOAA/NWS should adopt a core philosophy of molding the institution towards the Open WCS (Weather and Climate Services) paradigm whenever and wherever possible" (NOAA SAB EISWG, 2011). In fall 2011, the EISWG presented a report to the SAB titled "Towards Open Weather and Climate Services" that

was approved and transmitted by the SAB to the NOAA Administrator in late December of that year (NOAA SAB EISWG, 2011). The report finds that:

> "[T]he Nation has yet to realize the full value of NOAA's weather and climate services for two reasons:
>
> - First, various barriers inhibit the ability of NOAA to distribute or otherwise make available all of its weather and climate information, particularly high-resolution datasets such as numerical weather prediction model output, satellite and radar data.
> - Second, new technology and services are not developed within NOAA in a sufficiently symbiotic manner with the broader community such that optimized value from that new service or technology to society is quickly realized." (NOAA SAB EISWG, 2011)

In short, a working group with significant industry participation issued a report that recommended increased access to NOAA data for commercial users, along with their direct participation in the design and development of systems to collect and deliver it. The effort that began with *Fair Weather*, then, which spawned both a revised partnership policy creating avenues for influence by the private sector, and a committee to exploit them, now culminates in a report that recommends a significantly greater level of technical and decision-making integration with NOAA to address commercial needs. The following discussion highlights the main "enhancements" requested in the report, casting them in the light of the larger paradigm of commercial influence on NOAA data and decision-making.

Increasing Access to Commercially Valuable Data. While the report refers broadly to Weather and Climate Services (WCS)-related information available across NOAA, its primary focus is on those data available through the NWS. The underlying assumption is stated early on: "[O]nly a tiny fraction of all NOAA information is actually made available by the NWS for use outside of the agency (or even elsewhere in the agency in many cases)" (NOAA SAB EISWG, 2011). The authors emphasize that this is not due to "any systematic censoring policy," but to the large and increasing volume of data being created and the difficulty in communicating it due to technical limitations. The following figure excerpted from the report demonstrates their view of this barrier:

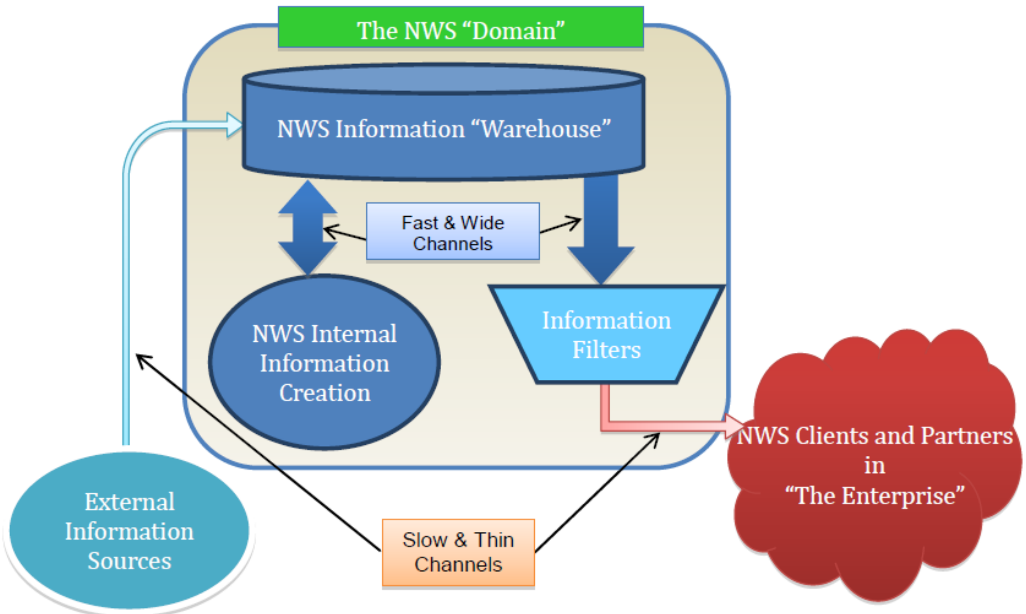


Figure 1 - NWS Information Flow Model (SAB, 2011)

The data that the NWS collects and creates ("NWS Information Warehouse") exists in an internal computing infrastructure where NOAA computing models (and resulting outputs) are directly connected to it via "fast and wide channels." The core argument supported by the diagram is that the "slow and thin" channels for access available to commercial parties (the cloud labeled "NWS Clients and Partners in "The Enterprise") prevent them from accessing large amounts of data in real-time. Due to these limitations (for technical reasons, but also others addressed below), the NWS engages in "information filtering" which limits the data (and potential value creation) available to commercial users. A table taken directly from the report provides some examples:

- NWP forecast data is filtered, particularly in time. Forecast models produce forecasts in time increments of several minutes, but only a small subset (about 1 in 100 typically) are actually published. Depending on the model and publication portal, spatial filtering of the model data may also take place.
- NWP ensemble forecast data is further filtered by parameters, with typically a very small subset of the models' parameters available in the model published in the ensemble suite.
- Satellite imagery is cropped and sub-sampled before broadcast on NOAAPort. Full resolution satellite data is available to the Enterprise through direct readout equipment, but this requires relatively expensive satellite receiving and processing equipment by each client wishing to receive the information. Unfortunately, the NWS does not publish satellite data via the internet unlike much of the rest of its real-time weather information.
- ASOS surface weather observing sites take observations every minute. Only hourly samples of the data are available, unless a significant change in the weather occurs, in which case an observation at the time of change is also published. Here, "significant" is largely defined based on aviation interests, which of course does not serve all interests.
- Data from the NWS NEXRAD Doppler data is first distilled into Level 2 moments (reflectivity, velocity, etc.) before publishing. The raw (Level 1) data is not published.

Figure 2 - Examples of NOAA Information Filtering (SAB, 2011)

It is clear, then, from just these examples that commercial parties are well aware of valuable data inside NOAA that they cannot access. The November 2011 presentation to the SAB that accompanied the report provides some examples of products that could be created with greater access. While some of the terminology used above is likely foreign to the layperson, the following excerpt from that presentation brings home the point:

> **Probabilistic Turbulence Forecasts.** By having access to each member of an ensemble at its "full-resolution," commercial firms could compute expected turbulence for each, then analyze the "spread" to produce an estimate of turbulence probability that would have value to airlines in route planning. Without this access, this product cannot be produced (by organizations outside the NWS). (NOAA SAB EISWG, 2011)

While the commercial value of this particular type of data is not explicitly stated in the SAB report, the 2017 Annual Research Applications Laboratory Report from the University Corporation for Atmospheric Research (UCAR, 2017) states "According to some estimates, turbulence encounters account for well over 75% of all weather-related injuries on commercial aircraft and amount to at least $200M annually in costs due passenger and crew injuries and aircraft damage." And, as a NOAA executive said in reference to the financial model underlying the NOAA Big Data Project (discussed in the next chapter): "All the airlines in the world might be really interested in buying those products. Yeah, so that would be one huge anchor tenant, basically an anchor industry, making that data available" (PRI-15).

Toward Symbiosis: Increased Participation in the Design and Development of Technology that Collects and Delivers Data of Commercial Value. The second barrier cited in the report is the lack of a "symbiotic" process between external users of the technologies and NOAA / NWS in their design and development. As context, these technologies vary in their uses, but the commercial sector's interest must ultimately be in those that collect, process, and deliver data. The reasoning presented, in brief, is that it reduces industry's learning curve when new or modified technology is implemented. This change would allow them to better anticipate changes in standards or data formatting, ensuring that "optimized value from that new service or technology to society is quickly realized" (NOAA SAB EISWG, 2011). The report provides detailed explanation and examples, but the core idea here is simple: By letting a company that depends on the end product become involved from start to finish, it will have more influence to minimize disruption to its information supply chain, increase efficiency, and improve the value of its contribution to their product(s). Granting this request could affect project priorities, budgeting, and technology investments inside NOAA, along with the measurements that are – or are not – collected, all further entrenching the requirements of commercial members of "The Enterprise" discussed earlier in the day-to-day decisions of government. The focus here is not on the net benefit to society of such a dependence, but simply that we are observing the attempted exercise of influence by commercial parties on the operation of a well-established information source upon which much of their business model is based.

The "Net" Effect. Even if government chooses not to let industry into the development process, the impact can come in a different form. Direct access to unfiltered data may result in identification of needed modifications in its content or structure to enhance its usability and commercial value. Witness an example given by one of the principals involved in the report: He points out that the structure of the processing that underlies numerical weather prediction models is such that results (data) are produced in incremental intermediate steps that run continuously every few minutes. However, the computer programming model that drives this process actually discards this data in the process of continuous probability development because NOAA assumed that these incremental results could never be made available externally due to limits on the capacity and speed of existing distribution channels (Neilley, P., Kyger, B., & Ramamurthy, M., 2014). So, he continued, it might very well be that the enhancement of access and distribution alone could result in "*changes to the model itself* (emphasis added) in order to store the data (now discarded) in such a way that it may be transported to the network near processing center" (run by the commercial sector) (Neilley, P., Kyger, B., & Ramamurthy, M., 2014).

Perhaps due to the arcane technical nature of the subject, it is easy to miss that this statement has a much larger and more generalizable implication for commercial influence in decision making about the content, processing, and distribution of government data: Intermediate steps in the construction of "data" for government use in its operations, even in much less complex situations, may have value to other stakeholders outside the local environment in which it is assembled. The idea of anticipated use as a factor shaping the construction of data addressed in the Introduction to this study is directly applicable here. From this, one can hypothesize that as commercial users take interest in data of all types, the changes and "time steps" that are summarized or discarded in its construction may be identified as commercially valuable. While in the domain of weather these incremental computations may not seem of dramatic import to the layman, as stated above, they are very likely the basis for new products worth millions of dollars. If this is true here, what other interim inputs into the production of facts across government may, when quantified and electronically captured, have value to new stakeholders, and, in turn, bring the influence of those stakeholders into the internal process of government decision making about data in other areas?

Action First, Policy Later. The report acknowledges that there are challenges to be overcome (security, cost, the extra burden of including the private sector participation in design and development), but also strongly, and repeatedly, qualifies those concerns by urging quick action – without providing specific answers as to how this might happen. Two aspects of the discussion are particularly relevant to this study's focus on the methods and impact of influence.

The first is that the unknowns involved are coupled with urgency, resulting in the SAB's recommendation to forge ahead, letting results drive policy. This echoes the use of subjective boundaries left open to industry for definition as a tool discussed in an earlier section, rooted in *Fair Weather*'s emphasis on "avoiding detailed and rigid boundaries" (NRC, 2003). Witness the statement from the SAB report that:

> Although there are numerous approaches that NOAA could undertake in implementing Open WCS (Weather and Climate Services), *it is believed that first developing comprehensive policies and procedures will not be effective* [emphasis added]. Such a process would be lengthy and likely not well-informed regarding the numerous nuances and challenges faced by the paradigm. (SAB, 2011)

Instead, the report advocates for incremental implementation on projects that "can be implemented quickly and will yield substantial benefits to the Enterprise, while also informing NOAA on a more comprehensive Open WCS approach" (NOAA SAB EISWG, 2011).

The second key challenge identified by the SAB in the report is “Fair Access.” We have already seen equity addressed in the NAO 216-112 partnership policy above. In this iteration, however, some of the conflicts latent in equity of access among commercial concerns begin to take shape – in this case 1) The challenges of integrating a large number of external participants in the design / development process when access “may not easily scale”; and, 2) The cost burden of access, with the SAB report noting that “since higher-volume data services generally come with more expensive access costs for the client, and since much of the information made available under an Open WCS may be high-volume, *the gap between those that can and cannot afford to participate may widen*” [emphasis added]. The report adds that, as a result “approaches and practices that minimize this potential segregation will need to be considered” (NOAA SAB EISWG, 2011). That said, of the nine potential projects offered by NOAA in their response to the SAB report outlined below, none appeared to involve the real-time access industry sought that would force this equity issue. But, as shall see later, this is not the end of the story.

Accommodating Commercial Interests: NOAA’s Response to the Open EIS Report. The normal operating procedure is for the SAB to address their report to NOAA and then, within a year, NOAA will issue a written response (Lubchenco, 2012). While I have omitted significant detail from both the SAB report and NOAA’s response due to space limitations, several situations addressed in NOAA’s response bear further discussion because they reveal fundamental conflicts or themes likely to be encountered in other government domains with regard to commercial influence and the development of mutual dependence.

NOAA’s Cost Avoidance Incentive**.** NOAA’s response stated that it “[W]elcomes the opportunity to further enhance this symbiotic relationship with an eye towards increasing value to the Nation in an economic environment where funding for new NOAA initiatives is challenging" (Lubchenco, 2012). Later, it also "…welcomes (the question of) how these policies can be more effective, especially in light of the fiscal challenges facing our country and the global environmental community" (Lubchenco, 2012). And, when the SAB suggests that NOAA "target accelerated implementation in of the Open WCS in specific areas that have limited risk or cost and can be achieved without a more comprehensive approach," (NOAA SAB EISWG, 2011), NOAA responds that "In this challenging budget environment, opportunities within current resources should be considered" (Lubchenco, 2012). These statements all reinforce the idea that one benefit NOAA receives from accommodating commercial interests is the possible accomplishment of some part of their mission for which they cannot, or are not required, to pay. That said, the language used here sounds like a tepid, rather than enthusiastic embrace of the overture.

Limitation by Financial Controls**.** Further emphasizing their financial limitations, NOAA stated in their response that they "must comply with numerous financial controls regarding the manner in which it (NOAA) acquires funds and the purposes for which these funds are used" (Lubchenco, 2012). So, if there are expenses involved in responding to industry demands, or NOAA were to "accept funds outside the appropriations process," they must have explicit authority to do so. This constraint is likely common to every government agency. Simply put, optimizing data collection and delivery for secondary users may cost money that is not budgeted. One exception that foreshadows the approach used later in the Big Data Project is included in a footnote:

> Cooperative Research and Development Agreements (CRADA) are examples of a mechanism for external parties to obtain proprietary access and, if needed, to pay for the cost of such access. Such agreements, however, will always be evaluated by NOAA from a broad public interest perspective. CRADAs should be viewed as a temporary mechanism to gain enhanced access to NOAA data and expertise." (Lubchenco, 2012)

***Barriers to Extension of the Commercial Information Supply Chain into NOAA*.** NOAA presents four significant, if not insurmountable, difficulties in responding to the request to extend the commercial information supply chain to gain direct access to data for real-time processing. The first is "**Limitations on the internal architecture of NOAA's data systems.**" The language used to describe these limitations evokes an image of a technical and information environment ill-prepared to support a reliable supply chain of high-resolution data to the private sector. They describe their data dissemination architecture as "disjointed and unable to transport all the data to a single portal for ease of data sharing" - adding in a footnote that the architecture is so fragmented that "there is no diagram available that accurately describes (it)" and emphasizes that – unlike the figure in the Open WCS report, "high-bandwidth connections *within* NOAA are more the exception than the rule" (Lubchenco, 2012). It also notes that "numerous data and resampled to lower spatial and temporal resolutions…even to users *inside* NOAA." Raw data from satellites "are even discarded in favor of the more processed, value-added version." The argument here is that NOAA's own architecture is not sophisticated or reliable enough to support the technology needed for real-time processing by the commercial sector. The second and third limitations are paired as "**Limitations of NOAA's capacity to provide data to external parties and limitations to the use of NOAA data by external parties after they acquire it."** NOAA's response here is straightforward: "NOAA doesn't currently have the bandwidth or potential server capacity to deliver the volume of data that might be needed by the private sector and academia" and continues "heroic (and expensive) system-level modifications would be required to make "purely internal, computer-to-computer intermediate results" available – with the ongoing costs of the infrastructure needed to do so "considerable" (Lubchenco, 2012). Perhaps stating the obvious, but this makes plain the case that NOAA - and likely most government organizations – did not design their collection systems to provide data beyond the audience of primary users, nor have they made the investment to develop and implement delivery systems to facilitate real-time, high-speed use of the data by external parties. This focus on the needs of primary, internal users by NOAA is the origin of the next limitation, where external organizations find themselves required to understand and translate "internal formats and encodings" (Lubchenco, 2012) to make use of the data, a barrier to efficiency, if not basic use. The internal practices, assumptions, and organizational methods applicable to data creation, including geographical reference points, may vary with the intended use, and may also be little, or poorly documented given that their use by external parties may never have been anticipated. Thus, conflicts between the data requirements of external users and internal users a) exist, b) modifications in favor of external users may impact internal uses of the same data by affecting standards, retention, priorities, and system design and operation, and c) the documentation created for internal use may not serve private sector needs in understanding and using the data it pertains to.

The final limitation in the NOAA response that is significant for the purposes of this study is a characterized as a "policy challenge" (Lubchenco, 2012) involving **requirements for Information Quality (IQ)**. The issue at hand is the federal Data Quality Act (Public Law 106-554) and NOAA's obligation to complete a determination as to the data's "quality, objectivity, utility and integrity" before releasing it (Lubchenco, 2012). As an example, they cite the case of "intermediate" satellite data and the need to distinguish between a release of this uncalibrated data versus NOAA's "finished products." Jeopardizing perceived quality of government scientific data related to meteorology looms larger here when one considers the political conflict around climate change, for example, as well as the public's ability to rely on its quality.

These limitations may seem very technical or abstract. But, if they exist with seemingly objective and resolute scientific data in a large scale computing environment, one might wonder how they would translate to measures that are easier to grasp, say, coding of local violations by the police, or the data and categorization used for social service clients or services, or – as has already been demonstrated in the last chapter, local restaurant inspections – all measures of external interest. One outcome, then, depending on its value to industry and their influence, might be that agencies are pressured to make these investments and redesign collection, distribution systems, or the data itself to make it reliable and of sufficient quality for commercial purposes. On the other hand, these barriers may be insurmountable from a financial and technical standpoint. As we shall see in the next chapter on the Big Data Project, the jury is still out.

Moving Forward

NOAA's response to the Science Advisory Board was dated October 24, 2012. It contained a timeline that proposed a list of nine projects / prototypes for consideration, committing to make a final selection and move forward in March 2013. Six projects addressed the joint (symbiotic) development concept and the three remaining projects proposed somewhat more direct and frequent access to NOAA data. Yet, none appeared to include the type of real-time access envisioned in the SAB recommendations. Undeterred, however, the commercial sector moved forward almost immediately with a project to obtain real-time access to numerical weather prediction model data by placing computer servers inside leased commercial space adjacent to key NOAA computing equipment that could be connected by a high-speed line. This occurred with the knowledge of, but without formal sign-off from, the EISWG and NOAA. The following section summarizes the key aspects of the evolution of this initiative, one that provides insight into barriers to commercial influence and also sets the stage for the story of the NOAA

Big Data Project addressed in the chapters that follow.

Crossing the Line: Privileged Access and the Reston Project

The December 2014 Report to the American Meteorological Society's Commission on the Weather, Water, and Climate Enterprise Executive Committee contained a status report of the year's activities of its subcommittee that paralleled the purpose of the SAB's EISWG. That subcommittee is called the Committee on Open Environmental Information Services (COEIS) and its report included the following statement (Louis Uccellini is the director of the NWS):

> In early June, the NOAA Science Advisory Board's Environmental Information Services Working Group (EISWG) members were informed by the EISWG Co-Chairs (passing information along from Louis Uccellini) that the NOAA attorneys had cancelled the efforts to establish a CRADA with The Weather Companies for the purpose of establishing a data portal trial at NCEP that would have provided open access to much of NWS high-res model data. This was to have been a tangible step forward in response to EISWG's Open Weather and Climate Services (subsequently referred to as Open EIS) recommendations. Upon receiving complaints from several companies, NOAA concurred that to establish a CRADA would have provided TWC with "privileged access." (AMS CCWE, 2014)

The SAB report had, of course, recommended direct access to NOAA data and provided specific examples of its commercial value. And NOAA, despite caveats, clearly expressed support for the Open EIS concept and for the idea that experimentation could help inform policy. Taking their cue from this, industry moved aggressively to set up a pilot of direct real-time access to data through use of a Cooperative Research and Development Agreement (CRADA), a contracting agreement mentioned in NOAA's response, only – as we have just seen - to be blocked by NOAA after a year of preparation. This decision and the surrounding context is significant because it shows that extending the information supply chain into government brings the competitive forces of the marketplace along with it. One private sector executive directly involved in the project pointed out that the main complaint was not by an organization that had been left out, but another participant (PRI-11). In discussing the decision, a NOAA official involved in the situation, summed things up as follows:

> The Weather Company wanted to come in and plug directly into the "quote" database, if you will, and then be the conduit to share that information with everyone else. Obviously, the AccuWeathers of the world and all the other entities did not want to then have to rely on The Weather Company to get them the information. It came down to this preferability factor…That's where we could not give preferential treatment to The Weather Company and then have them potentially start saying, "Well, we value-added to the data and now we're going to charge you for the data." That's really where the crux of the issue came into play. (PRI-12)

Of course, the SAB had already anticipated the issue of competition between vendors in its report in the section on "Fair Access." However, we see here that it is more than unequal resources among vendors, but also the design of distribution that matters –again, important background for understanding the impact of this aborted attempt on the Big Data Project, which continued the negotiation of these issues of cost, equity, and vendor access.

Despite the implication of the statement above, meeting minutes and a related conference presentation (Neilley, P., Kyger, B., & Ramamurthy, M., 2014) confirm the involvement of multiple companies, including Northrop Grumman, Unidata, Unisys, Vaisala, and Weather Decision Technologies in the project along with The Weather Company (TWC) (AMS CCWE, 2014b). A description of purpose of the purpose of the project by Unidata (2014), a planned participant, states it was to create a "community operated experimental prototype….to receive unique, high-volume, Numerical Weather Prediction (NWP) data from the National Centers for Environmental Prediction (NCEP) supercomputers in Reston, VA sufficient to enable new value-added processing by the community that was not otherwise possible." This would require super high-speed connectivity of 10GB/sec, which could only be achieved by placing servers very close to or inside the NCEP data center (Neilley, P., Kyger, B., & Ramamurthy, M., 2014). At one point, Amazon Web Services and Google became involved, but "…the latter had suggested that they host the data center within their ecosystem. The upside is that it would provide unlimited participation, but the challenge would be in transporting data over a longer distance" (Neilley, P., Kyger, B., & Ramamurthy, M., 2014). From the standpoint of equity, the leaders envisioned that once the value of particular sets of data had been confirmed, "Community member organizations could then choose to establish a presence at the data center, selecting the model data of interest and either transmit the selected data of interest *en masse* [emphasis in the original] to other sites or preferably, processing the selected data locally before transmission" (Unidata, 2014).

Before leaving this discussion, several aspects of the structure of the CRADA being built to govern this new access model should be noted, as they can help set the table for consideration of choices made only a year or so later in developing a CRADA for the NOAA Big Data Project. First, in this case the CRADA was to be established between NOAA and TWC. However, as their principal executive characterized it, it would require TWC to execute it in a fair and open manner, allowing "anyone from the community to come in and participate within the practical limits that we only have a certain amount of space we can do this with" (Neilley, P., Kyger, B., & Ramamurthy, M., 2014). The partners would share the cost of the basic infrastructure, but be required to install and pay for their own processing infrastructure, application, and connectivity back to their organizations. In describing the end goal, the executive suggested that one outcome might be that a cloud provider, such as Amazon - or a set of them - position itself near NCEP in a similar manner and then companies could purchase service from those providers which would come with access to this data (Neilley, P., Kyger, B., & Ramamurthy, M., 2014). Finally, as if on cue, the tiered approach to equity again raised its head in the questions from the audience after the 2014 AMS presentation on the project: Q: "You had said that the public would be able to connect to pull data – will you have secure computers for them to use?" A: "When I said "public," what I really meant is it's open to sophisticated users of our community, not Sally out on the street there…" He continued that the participants saw it as similar to WSRD-88 (NEXRAD Weather Radar) data was distributed currently in the United States, where there were three non-profit entities with preferred access who were required to redistribute the data to anyone on a cost-recovery basis. "It doesn't look quite the same because process capabilities would need to be sold, but it still is a basic model that I think could work here" (Neilley, P., Kyger, B., & Ramamurthy, M., 2014). We will learn more about that project and its history in the final chapter of this study. However, several months later, the work on the Reston project was stopped by NOAA (Unidata, 2014) and the focus moved to a new NOAA initiative, the Big Data Project, where the story continues in Chapter 4.

From Service to Sales(force)^{tm}: Incentives, Framing, and Emerging Practice in Accommodation of Commercial Interests in Constructing NOAA Information

Up to this point, our investigation has focused on how commercial influences work "outside-in" through advisory groups and lobbying, to attempt to shape NOAA policies, data, and delivery methods to industry advantage. We now turn to an "inside-out" view of recent activities of the Customer Engagement Branch of the National Center for Environmental Information, the office of the NOAA Chief Economist, and processes used in determining priorities that are performed by NOAA's Technology Planning and Integration for Observation division (TPIO). Each of these areas of the organization are enmeshed in a web of internal relationships with sub-organizations that are also subject to forces from external users of NOAA data. However, by isolating aspects of the work of select functions within NOAA focused on promoting, prioritizing, or assessing the value of NOAA information, we can learn more about how internal forces can intersect with and / or complement the outside-in approach to influence.

Engaging Industry at the National Centers for Environmental Information

The National Centers for Environmental Information (NCEI) serves as the primary archive and provider of NOAA data to external parties (Data Access, 2018) and considers itself the “"Nation's Scorekeeper" regarding the trends and anomalies of weather and climate” (About Us, 2018). It is a part of the National Environmental Satellite, Data, and Information Service (NESDIS), one of six major “line offices” of NOAA (NOAA Organization, 2018). NCEI is made up of four centers located around the United States – the one in Asheville, North Carolina houses its headquarters - along with six regional climate centers and four cooperative institutes. Regional climate centers are hybrid federal-state organizations primarily focused on providing environmental information specific to a specific geographic region of the United States (Climate Centers, 2018). Cooperative institutes are research organizations, usually associated with universities that perform research in collaboration with NOAA (Cooperative Institutes, 2018). The NCEI Customer Engagement Branch is an organizational unit that sits at the “tip” of the NOAA information supply chain as it makes its way out of the NCEI archive to external users. They are a primary interface with industry for its acquisition of NOAA data. Because of its high interaction with industry and the public (as many as 15,000 inquiries per year) (Dissen & Brewer, 2018), it is an ideal place to learn more about not only industry dependence on NOAA data but also about how NOAA comes to depend on and promote industry interests. We begin with a collaboration by NCEI that involves one of these cooperative institutes, the Cooperative Institute for Climate Studies – North Carolina (CICS-NC) via a contract with NOAA.

Origins of Enhanced NCEI Engagement Efforts

During the 2000's, a proactive strategy began to evolve (CICS-NC, 2012) that, by the end of the decade, led NCEI to begin reaching out, in partnership with CICS-NC, to users of climate and weather data in a focused "engagement" program of increasing sophistication (CICS-NC, 2018). This strategy of engagement appears to have had several sources: 1) The continuing advisory efforts that began with *Fair Weather* (NRC, 2003), then the Partnership Policy and SAB Open EIS report, as well as other National Research Council reports (NRC, 2001; NRC, 2009; NRC, 2012) that focused on opening up access to NOAA information of value to industry, to the point of recommending "identify(ing) new or evolved NWS data and services that can enhance public value delivered through the secondary value-chain" (NRC, 2012), 2) A large and visible effort in the federal government to address climate change, culminating in a NOAA strategic document *A Vision for Climate Services in NOAA* (Solomon, et. al, 2009) was coupled with proposed legislation that, while unsuccessful, would have created a Climate Service within NOAA; and, 3) NOAA's ongoing mandate to provide information about climate as part of its programs – say, in support of fisheries (Mandates & Drivers, 2018), drought monitoring, or coral reef management (Koblinsky, Pulwarty, & Davidson, 2009).

The purpose of the proposed climate service in the NOAA vision (Solomon, et. al, 2009) was:

> [T]o provide credible and authoritative climate information that will assist the nation, and by extension, the world, in developing and evaluating policy options for climate change mitigation and (will) enable decision makers, including resource managers and the public, to better anticipate, plan for, and adapt to impacts of a changing climate.

In turn, the engagement concept become centered on the idea that businesses could make use of NOAA data to prepare for the impact of climate change on their products and services, increasing overall resilience - and on stimulating a new industry of services to assist businesses in “climate adaptation” that was poised for growth (Dissen & Houston, 2014). The stage was also set by a 2013 (since-rescinded) Executive Order 13653 *Preparing the United States for the Impact of Climate Change* that encouraged government to bring resources to bear to help society as a whole prepare. While this specific combination of forces may not be readily found in other domains of government, the idea of using data in support of a programmatic outcomes seems likely to be much more common, only. However, as we will see shortly, in this case the data gained about use serves not just to inform industry but becomes a resource of commercial value itself.

Goals and Methods: From Responsive to Intentional

In a 2015 presentation, a member of NCEI’s customer engagement section outlined the transition to a new strategy. In the past, their work had been program-specific, leading to “many useful, single-purpose, customer-driven products,” but “knowledge of the customer was locked away in a few heads.” (Brewer, 2015). A recently commissioned market analysis of the use of NOAA information in the Energy Sector (2015) had produced recommendations that informed the new strategy. Among the areas listed for improvement were proactive engagement with users – and sectors, developing a better understanding of users and “entrain(ing) engagement into center business” (Brewer, 2015). Here we see the beginnings of attempts to further develop – if looked at from the commercial perspective - a channel that could be used more effectively for industry influence on NOAA’s data practices. As one NCEI employee said “It was really during that time when the whole idea of ensuring that we are a ‘use-inspired’ agency, a use and science agency, was kind of coming to fruition” (PRI-13).

The example used in the presentation of a use-inspired record was data on the extent of snow covering the ground (used by shipping companies, energy traders, and others). And, the approach defined for this type of record was to "find users and incorporate feedback into all stages from design to implementation (and improvement)" (Brewer, 2015). This theme of use-inspired vs. basic research reappears in conversations with and presentations by NCEI. While the word "use" lends itself to reinforcing the integration of commercial data needs into NOAA business, this is a larger issue. As the effects of mutual dependence on information are studied further, there will certainly be other prevalent cultural values, norms, and professional standards that inform the business of decision-making (beyond law and regulation) in other government domains that should be considered in that research. In considering NOAA a test case from which to abstract general hypotheses, and given NOAA is primarily a scientific organization, it should be noted that this idea of "use inspired" research is a fundamental philosophical choice in research orientation (Stokes, D.E., 1997). Past research has shown that funding of an agency's research by a national defense agency, essentially use-inspired, could subtly influence the agenda for later basic research by setting its early direction, scope, and emphasis (Oreskes, 2003). And, in the wider community, there are debates about the "corporatization" (de Oliveira, M.B., 2013) or commercialization of science – not the use of the result, but how a use-inspired or applied approach may tend to internalize neoliberal market norms in decisions about <u>what</u> science is done (and, as a result, what data is obtained or created). It was manifested here by a comparison between what was characterized as a "build it and they will come" approach (a negative connotation) and "having the requirements up front, which is really use-inspired" by a person at NCEI (PRI-13). While I did not broach the subject directly with NCEI of how commercial influence entering the organization through the re-use of data might begin to affect decision-making about the subject of what research is performed in this grander sense, clearly the intent and connection is there. This potential impact (What shall we work on and how?) seems easily generalized to other domains of government, suggesting that professional norms and standards for problem definition and prioritization be examined for the impacts of influence, along with inherent conflicts between values, say of equity, associated with the domain and those that arrive with increasing commercial value and government dependence on benefits from the private sector use of the data. For example, this concern is similar to one expressed in the previous chapter about such influences on archival decisions and appraisal, only in that case, about the impact of market values on what data will be retained.

To return to the subject at hand, the five components of the engagement strategy adopted by NCEI all involve better understanding of industry use of data and, in turn, soliciting and incorporating industry needs back into NOAA decisions about data, its structure and delivery

- Documenting customer requirements across programs;
- Enabling capabilities in customer analytics, which includes understanding and analyzing customer information, needs, requirements, recommendations, and successes;
- Understanding customers better through targeted regional- and sector-based interaction;
- Strategically prioritizing activities to address customer requirements; and
- Informing NCEI science, research and data priorities

(Brewer, 2015)

This move from "responsive to intentional" (Figure 3) has a direct bearing on the research questions in this study. NOAA / NCEI is making an intentional effort to solicit industry feedback and incorporate this into decision-making process about data collection and delivery by collaborating with two other organizational units of NOAA, the Chief Economist and TPIO. From a discussion with one of the principals involved in this effort in NCEI, it appears this new

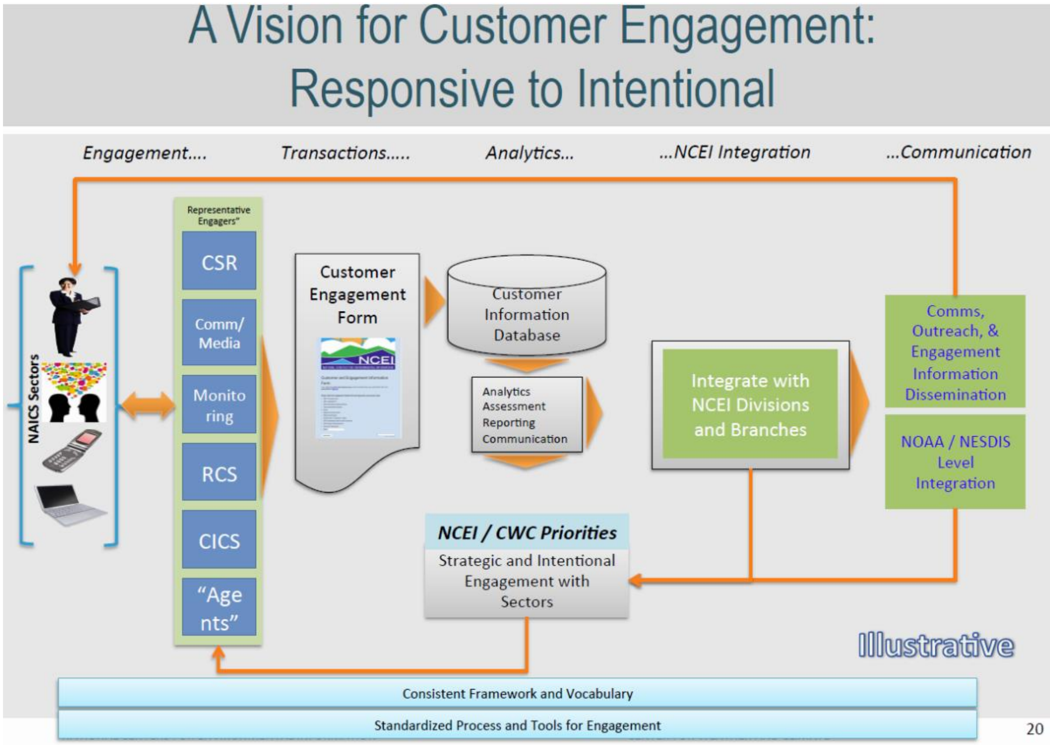


Figure 3 - A Vision for Customer Engagement (Brewer, 2015 – Backup Slide)

channel one of the principals involved in this effort in NCEI, it appears this new channel for solicitation of industry influence is emergent:

> So our idea is to take this from the product and service development side and say, "Look, here's the way that we do it. Here's the way that we capture our requirements. How can we come to a common way to capture both these service and product requirements and the satellite requirements that make sense to the people who are most senior in the organization?" We're actually working with the NOSIA folks (Note: NOAA Observing System Integrated Analysis standards – a methodology used by the TPIO for managing observing requirements that will be discussed later in this chapter). And if there's going to be new versions of those, how do we make sure that these aren't just for capturing satellite sensors and incorporate what we do and then we've got all of that information to go in and backfill. And potentially then jump on their structure and use that as ours instead of having our own (PRI-13).

He indicated they were also working with the Chief Economist…"[T]o the point where the kind of information we're putting together feeds directly into a more rigorous economic analysis that would be the purview of that shop" (PRI-13).

Aligning Mission to Market: Reconceptualizing Promotion, Policy and Priorities through the Lens of the North American Industrial Classification System (NAICS) Code. One technique employed as part the Customer Engagement strategy stands out. First, due to its similarity to a private sector approach (with accompanying values and logics) and, second, because it appears to be a new and unique case in the federal government in analyzing secondary use of information. As we have seen above, NOAA recognizes that many commercial parties depend on their information. A decision to use North American Industrial Classification System (NAICS) codes as a framework to drive "engagement" is an interesting development. While the coding system comes from the Census Bureau, which, like NOAA, is part of the Department of Commerce, we can see the beginnings of an attempt to market information products and gather requirements through a commercial lens that could drive future NOAA decisions about priorities, data collection, and even distribution. In the slide from an NCEI presentation below (Figure 4), for example, we see a map that casts industrial sectors that depend on NOAA data against their contribution to national Gross Domestic Product as a way that can be used to prioritize them.

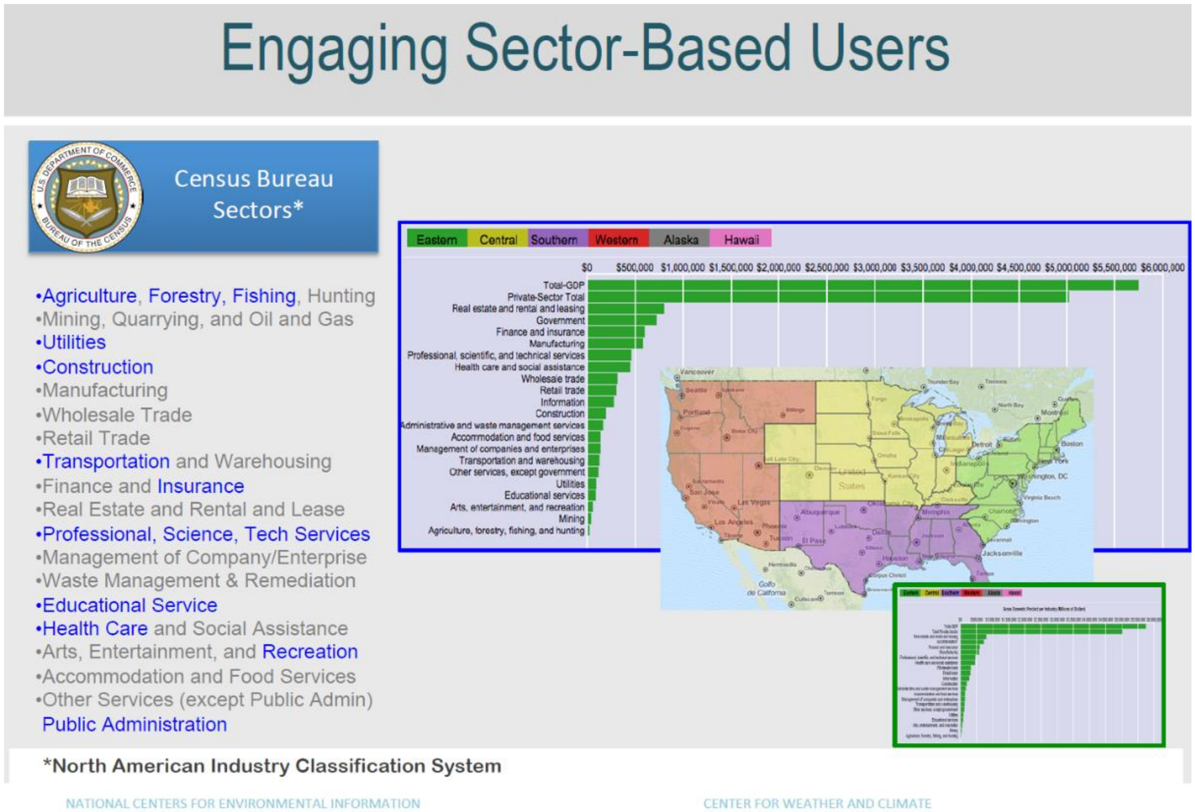


Figure 4 - Engaging Sector-Based Users (NCEI, 2015)

One of the principals portrayed this as an outgrowth of the need to prioritize their efforts due to limited resources (PRI-14). Making use of the industry sectors, they developed a prioritization system that started with asking if there were an existing mandate for the sector, or if something was required by law – in both cases, then, it was automatically “in.” Then, they considered the impact of the sector to the nation’s economy, and “What are the largest entities affecting our economy? Which sector has significant vulnerabilities to a changing climate?” Finally, they then assessed where NCEI / CICS-NC had expertise and substantial literature indicated extremes in climate were affecting the sector more (PRI-14). It follows, then, solicitation of requirements will not be approached in an egalitarian manner, but by prioritization that takes into account the economic value of the industrial sector. The result is portrayed below in a slide from the CICS-NC website (CICS=NC, 2018) that reflects a prioritization by industry sector in 2016 planning:

- **Prioritized Sectors for FY 2016:**

PRIORTIZED SECTORS	ACTIVITY	REGION	TIMING
Energy	• Energy Resilience Workshop Workshop • Utility Analytics Summit* • NOAA Energy Team • Industry sector engagement continues (Jenny Dissen)	• National / South • National	• January 2016 • April 2016 • FY 2016 - 2017
Agriculture / Forestry / Fishing / Hunting	• USDA, High Plains Regional Climate Center (HPRCC) and NOAA Regional Climate Services Meetings (Doug Kluck) • Ongoing APGA collaboration (Tami Houston)	• Western • Eastern • National	• March 2016 • March 2016 • Ongoing
Transportation / Infrastructure	• NC Department of Transportation Summer Workshop (Committee on Waste Management and Resource Efficiency) • ***Developing engagement options for FY17***	• National	• July 2016 • Focus for 2017
Government / Interagency	• NWS Climate Services Meeting • Ongoing Tribal Engagement • Department of Energy Partnership for Resilience • ***New opportunity with US EDA (in discussion)***	• National	• May 2016 • FY 2017 • Ongoing • Ongoing
Healthcare	• Partnership established with NIH / NIEHS • Working with NCA Climate and Health Team • Regional Meeting on Climate and Dengue Forecasting • ***Developing engagement options for FY17***	• State/National • National • International	• Ongoing • Ongoing • January 2016
Insurance	• NOAA's Strategic reinsurance industry engagement project (Adam Smith)	• National	• Ongoing
Finance	• ***New - developing engagement options for FY 2017***	• ***TBD***	• ***FY 2017***

Figure 5 - Prioritized Sectors for FY 2016 (CICS-NC, 2018)

To re-state, this approach reflects a view of data reuse and accompanying responsiveness seen through the value of the data to industry. And, it is reinforced by two other elements of the engagement strategy that provide data to further inform these priorities as we shall see in the following sections: Collecting data about NCEI customers and promoting NCEI data to business to jump start a climate adaptation services market in the United States.

Making the Commercial Sector Legible: Salesforce^{tm} Implementation. The first method used to inform the analysis by sector is the implementation, in collaboration with CICS-NC, of a state-of-the-art customer relationship management software package, Salesforce tm, to collect and analyze more information about the users of NOAA data. While implemented in fall 2017, the system already "[H]olds over 25,000 customer information entries with the ability to derive initial analytics based on customer type, and type of environmental data downloaded," categorized by industry sector (CICS-NC, 2017). From the standpoint of facilitating commercial influence, this is relevant as by categorizing users of agency open data and the data they use, they accomplish several things: 1) Abstract data that allows them to prioritize engagement efforts (as we saw in the earlier section); 2) Better understand use of NOAA data and to some degree infer its secondary economic impact - useful in prioritization, but also promotion of the value of NOAA programs and data to constituents, and 3) Enable solicitation and cataloging interests and requests by the commercial sector for new or modified information "products" that, by making them concrete and tying them to an industry sector, allows them to be more effectively championed upstream in NOAA for inclusion in planning and design of information products. This is especially true given the legitimacy conferred by this process on them as "use-inspired." The overall approach and connections are only now emerging - as a key participant said "And then, where NCEI would actually do something about those needs and translate them into actual requirements, we're just at the very early stages of that." (PRI-14 There is another objective, perhaps less intuitive, and maybe surprising. This is the idea to use the intelligence created by analyzing the use of this data about the Climate Services market to promote it. As one of the principals continues:

> Why don't we have solutions? All we do is talk about climate impacts? Where are the product and services-based market solutions? They (the private sector) are not only responsible, they are equally at the table and arguably more significant in enabling the change that we need. So, what does that mean? How do we as a government entity inspire - via information content - catalytic activities for innovation in the development of products and services that address this market? Really what I think we're doing is representing a pulse on the Climate Services market…because no one else can but us – and we have the history. I feel like we are literally sitting on a gold mine of information. Along with collecting new information about uses from the Big Data Project (see Chapter 4 and 5) and expanding the use of Salesforce[tm] across NOAA, they will have a much better picture of the market. (PRI-14)

In essence, then, while the data on users is being collected to enable greater responsiveness by NCEI / NOAA and promote the value of its work, it is also an attempt to quantify the climate services market in private sector terms so that opportunities can be more easily identified by new entrants. One can see where data is used and for what purpose, in what sectors (see above), setting the table for "catalytic activities" that are intended to foster businesses focused on climate change adaptation. "It's important because in the private sector context, you're enabling not so much best practices, but (serving) as a shepherd of that use case. And that enables some of the private sector folks to be able to create their own business case around it" (PRI-14. This is a novel and innovative strategy. While it may occur in other domains, it appears to be enabled, at least in part, by NOAA's dominant position as a data provider to industry.

It has also required a significant policy change. When the government enters into contracts with information users for a fee, it will, by nature, learn more about the use and users of the information. However, the users of open data are often anonymous. To collect more data with Salesforce^{tm}, NOAA was required to post a notice in the Federal Register (Privacy Act of 1974, 2017) and file a Privacy Impact Assessment (Privacy Impact Assessment, 2018) to change the Privacy Act Statement posted on their website, which now includes the statement "Contact information may be shared with a FedRamp-authorized Cloud service, SalesForce^{tm}, for generation of analytic reports regarding importance of products to particular user groups, and trends in how requests may change over time" (Privacy Act Statement, 2018). The Privacy Impact Assessment filing sheds light on other ambitions "NCEI is trying to get meaningful information such as which products are important to a particular group of users or what particular variables within products customers from various sectors are asking for (ex. temperature, precipitation, irradiance)…If possible, we would also like to capture benefits that users derive from the data" (Privacy Impact Assessment, 2018). This provides an interesting challenge in that it appears that some of what is collected could be considered "Business Identifiable Information" (Privacy Threshold Analysis, 2018), including "records that reveal commercial operations" and exempt from automatic release under the Freedom of Information Act or other law (Safeguarding Information, 2018). A similar situation may exist at the state level with regard to records of this type. Again, competition may manifest itself in information policy. While I did not discuss this with them, such restrictions could potentially limit the "resolution" of the climate services market model that can be shared publicly in support of NCEI's climate adaptation service advocacy. Finally, it should be noted that the interests described here fit with the development of cost-benefit analyses for NOAA data collection programs discussed in the section about the work of the TPIO below. While these efforts are not yet tied across NOAA, one can see connections are emerging that would support greater commercial influence in NOAA decision-making processes: In fact, from these interviews, this is their stated and specific intent.

Promoting Commercial Dependence on NOAA Data: Success Stories on User Engagement. Another key part of NCEI's outreach strategy involved working through CICS-NC and a European consulting firm, Acclimatise (via a U.S. vendor) to commission a series of reports, supported by posters and short videos, that showcase the users and uses of NOAA data and its benefits (GST, 2016b). Eight reports have been issued from 2015 to-date (mid-2018) following a preliminary study of the Power sector. Branded "Success Stories on User Engagement" (Success Stories, 2018), the studies use primarily qualitative interviews and desk research to explain how NOAA data is used by key industrial sectors of the U.S. economy. In the process, they solicit ideas from commercial users about attributes that would make the data more valuable to them. Of course, one obvious danger of relying on paid research that creates

“Success Stories” as an empirically valid source of data is their potential to overstate the case for value or dependence. This is compounded by multiple difficulties in measuring value in the face of alternative sources or courses of action, or cases where the data is only one factor used by industry in decision making. For our purposes, however, and despite these caveats, a quick overview easily makes the case that there are large industrial sectors that depend heavily on NOAA data for their operations and that NOAA has a strategy to seek input to make its information “products” more useful to them.

While available space prevents an enumeration of the dozens of examples recounted in these reports, most, but not all, of the focus is on climate and weather data. Foremost, of course, is the weather industry itself, estimated in value at around $7B (GST, 2018a). Despite the obvious uses for both general and specialized weather forecasting, there are other, less obvious uses. A court can normally only accept meteorological data certified by NOAA as evidence – which, in turn, creates dependence by the legal profession on “forensic meteorology” to meet the needs of the insurance industry and others involved in litigation that has a weather-related aspect (Certification of Data, 2018; GST, 2018a). This idea of designating certain data the official record, usually held by government, and controlling aspects of its use and access to it may be generalizable to other situations in its ability to confer value. In addressing dependence in the transportation sector, the $60B rail industry correlates historical weather conditions with weather-related derailments (the majority of 10,000+ in the last decade) to predict and avoid costs, rerouting trains to avoid snow-blocked tracks, or sending patrols to inspect for “sun kinks” in the rails based on detection of large changes in temperature (GST, 2018b). The report notes that Accuweather provides services to all seven Class 1 railways. The $100B livestock and poultry industry and its related ecosystem of “ranchers, livestock prospectors and traders, livestock associations…" all depend NOAA's U.S. Drought Monitor for decision-making on where to graze animals, to thin or expand the herd, or investments in feed and grain (GST, 2017). Another report is devoted solely to the Reinsurance industry (GST, 2016), the "insurers of insurers" that distribute the cost of risks of catastrophic events like hurricanes globally. They employ "catastrophe modelers" who use NOAA data to help estimate and price these risks. A more mundane use of NOAA data is by retailers like Foot Locker (GST, 2017b) who use climate reports to understand and predict variation in budgeted utility costs, as well as correlate changes to sales, say, of apparel, or Kohler Engines and Honda (GST, 2017b) that use these reports related to lawnmower products. The latter makes use of the Drought Monitor discussed earlier as well, which shows conditions on a county-by-county basis. In many cases, this data is used for making stocking and distribution decisions – or in one case designing or modifying sales territories (GST, 2017b). An example is provided of an investment company that, in evaluating performance of a chain of auto repair businesses for acquisition, was able to use the Regional Snowfall Index to correlate lack of snowfall with less accidents, helping to refine their valuation of the business that appeared to be doing poorly, but had apparently just been the victim of a streak of good weather! (GST, 2017b)

These anecdotal excerpts from the reports represent existing demand and dependence. But, in a complementary effort, *the consultants are also collecting suggestions for how the data can be made more useful to industry*. The reports suggest that customers of the U.S. Drought Monitor would like more data collection points within counties and the integration of more satellite data in the system (GST, 2017). Railroads would like data in different formats that covers more current periods (GST, 2018b). Investment companies that study the retail industry would like the Regional Snow Index to be updated more frequently (GST, 2017b), whereas catastrophe modelers need hurricane data more frequently, but also have more complex requests to address inconsistencies and improve analysis efforts in two comparable hurricane tracking models (GST, 2016). While in and of themselves, this selection of requests may present nothing dramatic, as noted earlier, NOAA – through its "front door" call center at NCEI – received 15,000 inquiries for data per year (Dissen & Brewer, 2018) and makes available 66,000 or more datasets (NOAA Data Catalog, 2018). The diversity of needs and their financial impact present the circumstances that policy feedback theory suggests should bloom into influence. In the following section, we will see NCEI reach out to the Power Industry, another large user of environmental data for load forecasting for traditional (coal, hydrological) and alternative (solar, wind) power sources (GST, 2016b), to investigate modifications to "Climate Normals," one tool they use for accomplishing that work. NCEI clearly sees the potential for this work to benefit them through greater visibility and support for their programs and more interest in their goal of building a market for climate adaptation services. While it remains to be seen if it will work, we already see signs of the extent to which they are willing to reshape themselves to try.

Addressing the Information Needs of the Commercial Sector: Climate Normals. The commercial influence discussed in this chapter is, for the most part, embedded in policy mechanisms or just now being engaged through new, but concentrated efforts to solicit commercial requirements with the promise they might be integrated into NOAA planning and prioritization processes. However, here we explore a instance where NOAA itself convened representatives from the energy industry and other stakeholders to gain agreement on a supplementary data product for industry use in decision making and regulatory compliance in rate setting at the state level. While this step seems "incremental" – data is not replaced or modified, only supplemented – it provides food for thought about the interaction of the parties and the issues involved that may prompt similar, or more aggressive solutions as efforts to build both dependence and engagement increase related to other types of data.

The data in question is an indicator called the "30-year Climate Normal" published by NOAA. These Normals are measures of climate, like temperature and precipitation (Arguez, et. al, 2013). When their composition and frequency were established in the 1930's, however, they were not designed to measure climate change, as it was assumed that the average varied little over time (NCICS, 2012). However, commercial uses outside of NOAA grew and they are now used for purposes as diverse as crop selection and building design (NCICS, 2012). One significant industry sector that depends on the Climate Normals are energy companies who use it for forecasting loads and designing plant and equipment (Arguez, et. al, 2013). The 30-year Climate Normals are also required as part of their regulatory filings with state government utility oversight commissions. For NOAA's purposes, the Climate Normals have traditionally been updated once per decade (Arguez, et. al, 2013). However, commercial users in the energy sector indicate they need more frequent updates due to the impact of climate change on both forecasting and in justifying rates and investment with state utility commissions.

In an early initiative in the NCEI engagement efforts, they worked with CICS-NC to convene an Alternative Climate Normals workshop in spring 2012 that included 50 stakeholders from the energy industry, state regulatory bodies, and climate science to discuss possible alternatives for Climate Normals to meet their needs (NCICS, 2018). The outcome of the deliberations was that "NOAA is strongly encouraged to develop alternatives to its traditional climate normal." The findings also emphasized that these should not be labeled as "experimental" or "alternatives" (Arguez, et. al, 2013), but be part of the "official product suite," assumedly so that they would have the imprimatur as legitimate indicators that could be used for regulatory compliance. One can also see a thirst for more detailed, higher resolution data. An article summarizing the outcomes of the meeting states that industry was interested in obtaining 10-year hourly Climate Normals (Arguez, et. al, 2013), and that processes currently performed on a monthly basis to remove bias from changes in station location or citing, temperature instrumentation, and observing practice be applied on a daily (and hourly) basis. NOAA indicated they would move forward on these recommendations and currently produces *supplemental* (emphasis added - note the term used) Climate Normals that include these 10-year hourly measures.

The Financialization of Environmental Risk and Influence on NOAA Data

In the last two sections, we've seen how industry has come to depend on NOAA data for decision making, prediction, and managing risk. In each case, one can see the possibility that this dependence might result in attempts to influence the content, timeliness, and availability of data to improve business efficiency and performance. However, another aspect of NOAA data not directly addressed in the work of the NCEI has significant implications for influence. This is the case where as part of the "financialization" (Pike & Pollard, 2010) of the environment and societal / business risk, financial instruments like derivatives contracts come to depend on NOAA environmental indicators as part of their underlying risk model. The most visible example is the use of temperature and precipitation data in weather derivatives contracts (Randalls, 2010; Bates, 2014b) where insurance companies may base payouts solely on the value of an indicator of a condition in the physical environment (or set of them), rather than assessing damage or specific instances involving local conditions. Randalls (2010) notes situations where a weather derivatives contract payout may be triggered based on a small variation in a temperature measure of one or two degrees – causing a natural tension when the government decides to move, replace, or recalibrate equipment, as well as a rift in expectations about investment "Weather traders thus claim that more money should be devoted to generating accurate observations rather than developing better models or synthetic data: a plea for reorganizing meteorological priorities..." (Randalls, 2010).

As more environmental measures are developed and become available, the opportunities to quantify risk based on their variation and arbitrage it in financial markets increases, exposing a wider array of data to commercial dependence on its stability and consistency tied to the structure of the financial instruments that depend on it. Literature suggests growing use of these financial instruments (and associated measures) for fisheries (Little, et. al, 2014), environmental conservation (Lemes, La Nauze, and O'Neill, 2011; Sullivan, 2013), and via catastrophe bonds (Blackman, Maidenberg, & Varnham, 2018), financial instruments sold in place of traditional reinsurance to insure against natural disasters like earthquakes or hurricanes. While further analysis of the growth and frequency of the use of these indicators is beyond the scope of this study, in light of research already published in this area and nascent efforts to introduce natural capital accounting (see Chief Economist discussion in the following section) that would financially quantify stocks of natural resources and develop measures of the changes in their value, this area seems ripe for further research on the effects of the dependence of financial instruments and markets on these indicators. Equally important from the perspective of generalizing the results of this study to the consideration of other types of data, it appears that the logic used in these instruments could be extended to a broad array of socioeconomic indicators like crime rates (for real estate pricing), for example. Firms specializing in providing "alternative data" to the hedge fund industry for use in market models have found a wide variety of "leading" indicators of financial performance for use in these predictive models, from railroad traffic to time-series data on energy use from the U.S. Energy Administration (Quandl, 2018).

The Enterprise and "Societal Benefit": Integrating the Commercial Value of Information into NOAA Priorities

The story of supplemental Climate Normals is just one example of how NCEI solicited industry feedback in shaping decisions about data structure and availability. The "Success Stories" outreach documented dependence in other industrial sectors and also solicited similar types of input with the implication it would be considered in NOAA decision making. In the course of talking about that work, NCEI officials mentioned emerging relationships with two organizations inside NOAA: the NOAA Chief Economist and the Technology Planning and Integration for Observations (TPIO) division. This section briefly examines the work of these areas as it relates to NOAA's decisions about data and considers how their mission, and some aspects of how they help NOAA set priorities, might make them amenable to NCEI's proposals and result in a new conduit for industry to exercise influence on NOAA data, its structure, and delivery.

The Value of Information: NOAA Chief Economist. The NOAA Chief Economist resides in the Office of the Chief Financial Officer (OCFO) inside the Performance, Risk and Social Science Office. As part of efforts to integrate social science into the agency's work, one core function of the office is to "quantify and promote value and impact of NOAA's products and services" (Wilhelm, 2016). In support of that function, there have been efforts toward "developing a "Value of Information" approach and research to quantify the impact of NOAA's products and services" (Wilhelm, 2016). The office is also working on including impacts and value related to "ecosystem services" into NOAA decision making. This concept is defined as "the benefits that flow from nature to people" (OMB, 2015), and involves quantifying natural resources of all types as assets and incorporating the resulting impacts on those resources when considering alternatives. The concept is related to the emerging practice of “Natural Capital Accounting,” a global initiative to integrate the value of natural capital - “the earth’s stock of natural assets” (Natural Capital, 2018) – into planning, accounting, and cost-benefit analysis. This idea is significant for this study’s research questions as *it positions NOAA data as a tool to quantify environmental asset values* – a topic of great commercial interest. For example, the Chief Economist is also involved in an initiative of the international Group on Earth Observations (GEO) called “Earth Observations for Ecosystem Accounting” (GEO, 2018). This is an effort to develop a framework for using earth observations in “monitoring ecosystem extent and conditions” within this approach to accounting, something that could easily be rephrased as “environmental asset valuation monitoring” from a commercial point of view. The group charged with this would eventually develop “definitions, indicators, scale and temporality, sampling regimes” and other measures tied to particular earth observations (GEO, 2018). If the effort evolves successfully, it will more deeply intertwine the composition of data and priorities for NOAA’s investment in it given its use as an “independent” measure of asset condition, serving both governmental and commercial purposes. In other words, what’s measured and how, with consequent pressures to anticipate and accommodate its dual role as environmental measure and financial indicator. To the degree that commercial entities would depend on these measures to structure financial instruments, say, to insure a particular set of “environmental assets” and validate their financial reporting, this implies further influence and negotiation on the composition and structure of the measurements used for this purpose, along the same lines as earlier discussions above about the commercial dependence of financial instruments on the stability and predictability of the composition of NOAA environmental measurements.

Why do these seemingly arcane issues matter to research questions about commercial sector influence in the construction of government data and access to it? The OCFO office is directly involved in planning and budgeting for the agency and also sits on the NOAA Observing Systems Council (NOSC), a key decision-making body for investments in NOAA data and collection (observing) systems (NOSC, 2018). And, in interviews conducted with individuals in NCEI (emerging source of new requirements and conduit of influence from commercial entities), TPIO (manages the prioritization of requirements and systems as support for investment decisions), and with the Chief Economist (charged with valuing contributions and liaison to financial frameworks that incorporate those measures), all identify recent outreach among the parties that indicates the beginning of efforts to tie these processes together.

Influence and the Efficient Frontier: Determining Value in the Prioritization of Observing System Requirements. In an earlier section of this chapter, an NCEI manager explained that adopting the same tool NOAA used for tracking and prioritizing requirements across the agency would increase the chances that the requirements they collected from customers would be incorporated in NOAA priorities. That tool is the *NOAA Observing System Integrated Analysis* (NOSIA-II) methodology (NESDIS, 2016), a portfolio analysis framework managed by NOAA's Technology, Planning and Integration for Operations (TPIO) division inside NESDIS – the same organization in which the NCEI resides. A description of the technical mechanics of the NOSIA-II and the related software used to manage and calculate scenarios for planning purposes from the results is beyond the scope of the present discussion. However, a basic outline of the process quickly illustrates how the measures used to assign value in the model open the door for potential incorporation of industry influence and requirements in the process, a process that drives NOAA priorities for investment in data collection efforts.

NOAA divides its strategic planning (NOAA, 2010) into four long-term goals: *Climate Adaptation and Mitigation*; *Weather Ready Nation*; *Healthy Oceans*; and *Resilient Coastal Communities and Economies*. In turn, each is broken down into an overall total of 26 Mission Service Areas (MSA). An MSA is defined as "a NOAA core function that is focused on a specific environmental process, socioeconomic sector or activity to achieve societal outcomes aligned with NOAA's mission" (NESDIS, 2016). For example, *Weather-ready Nation* is a catch-all for the concept that "Society is prepared for and responds to weather-related events." Its supporting MSAs include objectives like "Reduced loss of life, property and disruption from high-impact events," as well as "Improved freshwater resource management." Another is "A more productive and efficient economy through environmental information relevant to key sectors of the economy." So, it is in the furtherance of MSAs like these that NOSIA-II is used to aggregate and rank the contributions of observations (data measurements) and their supporting observing systems (ground sensors, satellites, etc.). In a recent effort, hundreds of internal subject matter experts were surveyed, including virtual or on-site visits to 72 sites (NESDIS, 2016). Critical to our analysis is that at present, while priorities are based on societal benefit, <u>no external users / uses</u> are included in the data collected nor are they explicitly factored into the prioritization process. The language used in the 2016 NOSIA-II report frames this as a deficiency, pointing out: *NOSIA-II capability assesses all observing system impacts have upon a sample of NOAA products and services. If the primary user of a NOAA observing system is external to NOAA, then that observing system's total impact to the nation will be under- valued by the NOSIA-II capability* [emphasis added] (p. 19). A similar statement is listed first in an appendix describing the model's limitations (NESDIS, 2016).

The purpose of the NOSIA-II methodology is to allow observing systems to be managed as a portfolio, by "document(ing) the relationship between available observing systems and their impact on NOAA's diverse services and scientific objectives" (TPIO Project Spotlight, 2018). This helps them determine the tradeoffs and the most efficient "mix" of investments to meet the mission and maximize the associated benefits to society (TPIO Project Spotlight, 2018). This optimal mix has a technical name, the *Efficient Frontier* - a "graphical representation of combinations of assets or investment options that provide the highest possible performance or goal satisfaction for a range of cost or budget constraints" (NESDIS, 2016). As by NOAA's own admission, the model fails to include billions of dollars in societal benefits provided by the commercial sector through use of NOAA data, it opens the door for either NOAA or the commercial sector itself to make the argument that this prioritization effort and the resulting Efficient Frontier used to make investment decisions are fatally flawed and must be corrected by including the benefits provided by external users of the data…*along with their needs and requirements*. Given the degree to which the policy setting discussed earlier in this chapter was predicated on the common goals of a "Weather Enterprise" that includes the private sector, the same type of framing of needs and benefits seems quite possible here. It is, after all, the same people.

As we leave this chapter, there is an internal incentive related to prioritization that should also be noted as having the potential to lead toward incorporation of private sector benefits in NOAA decision making. In interviews with the TPIO, it was stated that programs like NEXRAD used their high ranking in the NOSIA evaluation to lobby for funding to continue the program (the useful life of the underlying technology has limits, and investment is needed). In turn, in discussing the internal reactions to the final prioritization, some individuals took issue with the low prioritization of their programs, saying: “I feel like this system should have been ranked higher than it was” (PRI-33). Given the drive for efficiency and the challenging budget times frequently mentioned by NOAA management, one could see a case where those with information products that had large societal benefit - but via commercial users who reused their data - might have cause to join in the push to incorporate those requirements (and benefits) in planning and funding decisions. And, given the model’s premise, this seems quite logical. As one member of the TPIO said, “The Holy Grail for us is really to link to economic benefits, so we can do a true cost-benefit analysis” (PRI-33). The pieces seem to be coming together in a way that would include the downstream benefits from the private sector – and the data requirements used to obtain them – in that calculation.

Conclusion

Two cases, both involving the same large government organization and the same industry. Yet, neither are monolithic in their interests or approaches. In the first case, industry takes the next step in a long history of drawing the line on NOAA competition that could hurt their bottom line by…removing the lines. By defining data production, research and development, and service delivery across academia, government, and the private sector as the “Weather Enterprise,” the endeavors of government are defined into a common effort where ultimate goals are shared and efforts are in the services of what is best and most efficient for all – without defining it. There is no need for rigid definitions, given the fast-paced change of technology, capabilities, and needs between these groups. As a result, these lines between public and private become what I call “boundaries of convenience,” movable by industry based on industry interest as they have the upper hand, both symbolically and, as things later develop, through advisory boards and legislative sponsors. Industry is able to gain concessions in the terms of government's data production and delivery, getting NOAA to agree, broadly stated, not to change anything without asking their opinion, and keeping the data flowing and reliable where industry already depends on it. The is an interesting situation with regard to policy – technically the only policy case among those I studied – because it is a demonstration of what might lie beyond the basic FOIA, open data, or open records policies that represent most of the information policy in government today. When things get down to a policy specifically focused on providing environmental data to industry, they are at the table, and when there is a lot of money involved (here billions of dollars), we see how things could come out. Their intentions are further revealed as the advisory committee they form asks for even more concessions, deeper access, and involvement in design. A tepid response and offer to engage in exploring the suggestion is enough for them to immediately move to set up shop literally next door to a NOAA data center with the idea of running a high-speed line directly into their numerical weather prediction data to access it real time. This is stopped, but only by another vendor who feels it might disadvantage them. This idea of "equal, but different" demonstrates a gap in information policy and the critical role of decisions about what I'll call the architecture of distribution of government data to commercial concerns. It is widely left up to discretion of the managers involved – we see that in the first three cases in this study – and, here, the vacuum makes for continuous negotiation, and conflict, with government struggling to remain responsive in the face of resource limitations and the forces of path dependence in the form of its own organizational needs and technological investments that may not be aligned with private sector needs for their product.

The second case is different. Here, NOAA sees a potential benefit in moving from "responsive to intentional." While not pronounced in the discussion, it is interesting to note that the origin of these efforts is at the archive – the same site of influence and negotiation as we saw in an earlier case at the state level. This time, however, the archive is much larger and diverse, although it is still subject to the same issue – the private sector is not interested in all of its data, but opportunistically depends on what it has come to find useful. The NCEI turns this around, though, recognizing that promoting the use of their data – to the point of creating public reports, graphics, and videos of its impacts on business and society – can only improve their support. While we don't see specific evidence this is the case – the promotional effort doesn't come with readily visible metrics, other than its level of activity – we see that it represents a new way, or at least a more purposeful way of thinking about their role. And, from this we see other behaviors – as part of their outreach, they solicit input into how their "use-inspired" information products can help business, and start to identify how they might embed that in NOAA decision making. They start to catalog the activity of their data users, using a state-of-the-art customer relationship management system used by Fortune 500 companies to keep track of customers, sales leads, and business opportunities. And, they start to map these groups of users against their economic impacts on the economy, tying this activity to a larger mission – to model and "catalyze" industry interest and investment in a new area: climate adaptation services. Again, while this effort is just in its early phases, it is hard to tell what the outcome will be. But, for this study, it shows that influence sometimes works "inside out," not just "outside in." If government sees a potential benefit in secondary use of their data, they may start to change their own behavior – and their data products - to obtain it.

Chapter 4: The NOAA Big Data Project: Part I – Building the Disruptive Data Imaginary

Introduction

In the first section of the last chapter, we saw how “feedback effects” (Pierson, 1993) from commercial beneficiaries of the federal government’s open data policies shaped NOAA’s policies on information access. As we left it, the “Reston Project” used recommendations based on these polices to assemble a consortium of vendors led by The Weather Company (TWC) in an attempt to gain real-time access to NOAA numerical weather prediction computer models. After months of work, this effort was abruptly cancelled by NOAA, citing the “privileged access” its design would have granted TWC (AMS CCWE, 2014). Then, we concluded with a look at an entrepreneurial effort that seemed almost the opposite. NOAA’s NCEI, in anticipation of gaining benefits from the private sector and shaping <u>its</u> use of their data, undertook a program to do just that, promoting their work, cataloging their private sector customers to design marketing programs, and even soliciting their requirements for potential inclusion in the design of their data.

In this chapter, we pick up that story as we trace the evolution of another project promoted by NOAA that began around the same time as the Reston project ended, the NOAA Big Data Partnership (later “Project” or BDP). This project also makes use of a “Cooperative Research and Development Agreement” (CRADA) to incentivize and assist five organizations, including four of the world’s largest technology companies, in identifying NOAA datasets with commercial value. The agreement NOAA designs anticipates that copies of these datasets would be transferred to each company. In turn, the companies could obtain revenue from fees they charge other commercial users to use their “compute” (powerful and easily expandable computer processing resources) to process the giant datasets in place, or make money in other ways from the data through innovative “value-added” services they could develop. These added revenues would allow them to subsidize the cost of storing and providing the raw data to the public for free, reducing the burden on NOAA to do so. In this case, we see NOAA willing to work with and around their own policy to change practices in the pursuit of benefits, designing an approach that industry didn’t ask for, but that they are sure they will warm up to once they agree to participate.

The explicit intent of the project, then, is to use a new policy vehicle to attract and assist industry in identifying and commercializing public data to increase its use (Industry financial benefit / NOAA mission benefit) while reducing NOAA’s costs to meet growing demands for access (NOAA financial benefit through cost avoidance). This formative moment (Pierson, 1993) is an ideal place to continue our study of the elements involved in developing mutual dependence and related influences on government data and their impact. Through interviews with current and former participants from NOAA and the academic and commercial sectors in this (still ongoing) project, along with a review of public presentations and supporting documentation about its goals and feedback provided by industry at the time of its conception, this chapter traces the evolution of the BDP from the conclusion of the Reston project to the completion of the new policy being “researched” in the form of a signed CRADA. It attempts to isolate the influences that shape the BDP’s origin and development and their impact on both commercial and government decision making about data and its distribution. In turn, we will see the implications of these choices play out in the following chapter where we present and analyze influences in the implementation phase of the project, its lessons learned (and ours), and future direction.

Prologue: Reston Revisited

We begin by looking more closely at the origin of the BDP, with the idea that the motivations that led to its creation may predicate the later growth of dependence and exercise of influence. There are signs that it could be an outgrowth of the Reston Project: It involves some of the same companies (The Weather Company, Amazon, and others) that will come to participate in the BDP and, at the time it was stopped, the participants were also developing a CRADA (Neilley, P., Kyger, B., & Ramamurthy, M., 2014; AMS CCWE, 2014b). The end of the Reston Project and the beginning of the BDP also occurred at roughly the same time. Did the BDP evolve from the Reston Project, perhaps driven by the same commercial forces that shaped earlier policy – or as an effort (real or symbolic) to assuage them? Or was it the result of other influences or ideas? It is important for the study of influence to understand who was influencing whom here.

Two Views of Reston

The Reston Project came to an end in June 2014 (AMS CCWE, 2014) while the first Request for Information (RFI) for the BDP was published in February of the same year (USDOC, 2014a). Documentary evidence about the cancellation or any relationship between the two projects was not presented, however, until six months later at the next regular meeting of the NOAA Science Advisory Board's Enterprise Information Services Working Group (EISWG) in December 2014 (NOAA SAB, 2009) where an "Open Environmental Systems Update" was on the agenda. At the meeting, NOAA distributed a handout with a section titled "Why the CRADA Fell Apart" (NOAA SAB, 2014b). They also made a presentation addressing both the Reston Project and the upcoming BDP, emphasizing their complementary nature. In turn, a presentation by the Reston project leader from The Weather Company (also co-chair of EISWG) included a slide titled “Why It Didn’t Happen” (NOAA SAB, 2014c) that emphasized a contrary view.

1) **NOAA.** NOAA gave the same explanation for its cancellation of the project noted in the previous chapter: “[I]t would have, in effect, established an environment that put The Weather Company (TWC) in control of the distribution of NOAA data,” violating both federal and NOAA’s information policy requiring equal access (NOAA SAB, 2014b). As a result, other companies were “demanding that NOAA not go forward with the CRADA” due to it providing an “unfair advantage” to TWC. However, the presentation stated another reason: “NOAA was also concerned that the effort to develop more comprehensive access to model data under a CRADA with TWC would interfere with development of the Big Data Partnership (BDP) solicitation, which will address all types of NOAA information (not just model data)” (NOAA SAB, 2014b). While this justification hardly seems as compelling as the threat of a company lawsuit against NOAA referenced in the summary below (PRI-11) its inclusion does appear to a) set the projects in opposition and b) implies greater NOAA support for the BDP.

   NOAA’s presentation also reiterated the problems with real-time access addressed in their formal response to the SAB/EISWG report, annotating the original diagram (see Chapter 3, Figure 1) on which the EISWG proposal was based, and displaying it as follows:

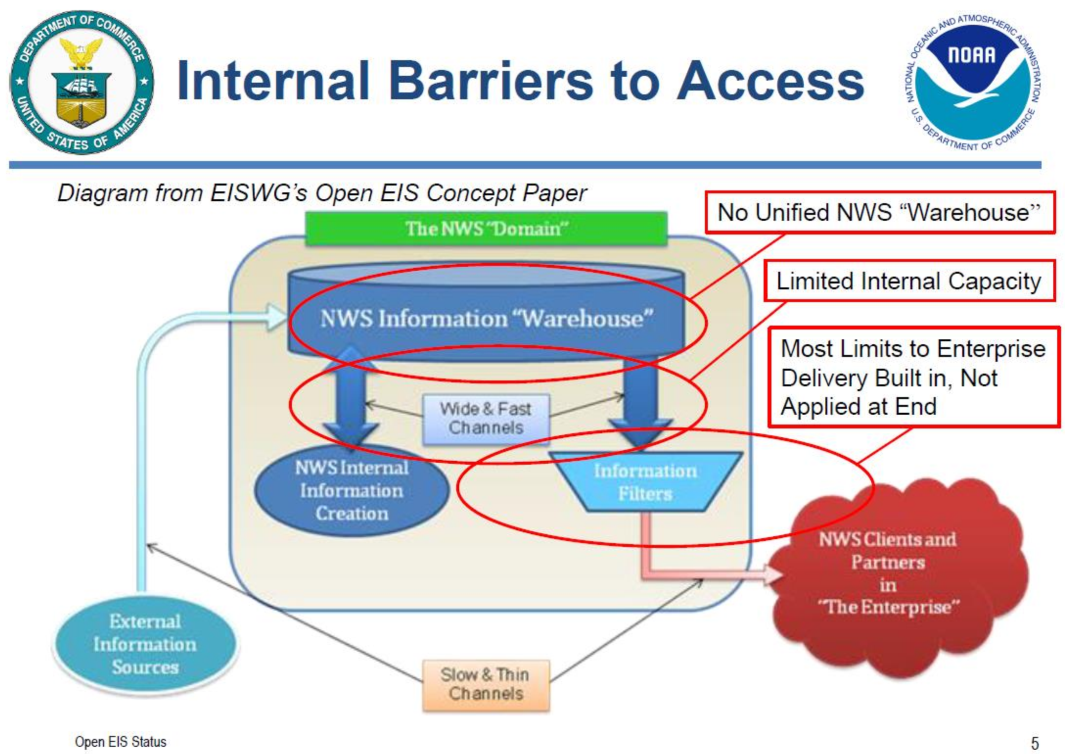


Figure 6 - Internal Barriers to Access (NOAA SAB, 2014d)

Essentially, the technical (rather than policy) issues NOAA had identified in their original response to the EISWG recommendations – see discussion in previous chapter (Lubchenko, 2012) - had not changed: The internal technical infrastructure was not sufficient to support the requirements of a commercial information supply chain for real-time access to data from National Center for Environmental Prediction numerical weather prediction (NWP) models. With these barriers and the policy challenge of maintaining equal access that had been evidenced in the Reston Project, NOAA now asserted control by laying out more formal constraints. The presentation stated that the National Weather Service (not industry) would decide what data would be provided experimentally (although an open call was made for suggestions), only existing (NOAA) infrastructure would be used (vs. the cancelled initiative that replicated data to commercial servers), and the NWS would "decide at some point in the future if this would be made operational" (NOAA SAB, 2014d). This restructured effort would be “Complementary to (the) Big Data Partnership, which will create new access paradigm(s)” (NOAA SAB, 2014d). *In NOAA’s official conceptualization, then, the BDP was a different project that would create a new access model(s).*

2) **EISWG / Industry.** Not surprisingly, **t**he perspective taken by the TWC representative from EISWG was different. A “Why it didn’t happen” slide begin with a bullet “The Partnership Policy: Fair to All?” and questions the validity of claims that its design was inequitable. It asserts that the data would have been available to all who participated and, while TWC would have led the program, it would have been managed as a “equal, equitable consortium” (NOAA SAB, 2014c). Given this, he posed the question “Is the Partnership Policy Too Rigid?” suggesting that perhaps “out” language should be added to enable NOAA to make decisions that are, on balance, right for the nation, but perhaps inconsistent with some aspects of that policy” (NOAA SAB, 2014c). Finally, in what seems like crucial evidence for a disconnect between the two projects from the perspective of industry interests, the TWC/EISWG presentation stated that the BDP did not appear to be a “symbiotic development project” nor “address access to new core model data,” nor “enable new methods to access the data such as stream computing methodologies” (NOAA SAB, 2014). In other words, his assessment was that the BDP did not meet any of the major objectives of the EISWG’s report detailed in Chapter 3, including the one the unofficial Reston Project had been attempting to move forward on, real-time access to NOAA Numerical Weather Prediction model data.

This episode reinforces a few important points:

- **The interests (and influence) of industry are not monolithic and sometimes conflict for competitive reasons.** In addition to the NOAA presentation above, participants also confirm, *it was pressure on NOAA from at least one or more vendors participating in the project that resulted in it being shut down* – in the face of competition, a vendor may work to block access if it would somehow advantage a competitor. A representative of one of the firms involved stated "They made us think they went to NOAA and said "If you let this go forward we're going to sue NOAA and you need to kill this the way it is….They actually propose something: "Well, we'll take it over" (PRI-11).

- **Equal but different – again: Equal access policy can be a cudgel, not just an enabler.** From the review of the NOAA Partnership Policy in the last chapter and its challenges with regard to equality, the barrier presented by the requirement of equal access across classes of users without regard to their resources and capabilities seems to be what is at issue here with "privileged access." Equality of access comes to mean something different when new technological methods require certain resources and capabilities to be exercised. In this case, NOAA asserted that – in essence - the design of the Reston Project failed to accommodate technological and resource inequity across interested parties – when one of the industry members strongly advocated that view.
- **NOAA has its own interests, among them security and their own technical capacity to support external supply chains, and is motivated to find a way to protect them.** We see this in the diagram above and their emphasis on regaining control to provide access on their own terms. The limits of this technical design evidence themselves in the next chapter on implementation.

In total, the answer to the question of Reston's part in the origin of the BDP is that 1) the BDP was not an explicit extension of that project, 2) It met none of Reston's (nor the EISWG's) goals according to a private sector participant, 3) Reston's goals were part of recommendations seen by NOAA as infeasible due to inadequate infrastructure, and 4) a continuation of that project might have interfered with the Big Data Partnership solicitation – assumedly by diluting the focus by industry on the BDP as the preferred NOAA approach which would push data out to public platforms, not allow industry to access data residing inside NOAA, preferences that reflected concerns about security and overrunning their infrastructure. So, where did this idea come from?

Origins of the BDP: "From Responsive to Intentional" Revisited

If the BDP was not created as a response to specific industry demands (at least those who had driven the unofficial Reston Project, nor the recommendations for greater access in the EISWG report discussed in Chapter 3) what was its motivation? Put differently, how could a project to facilitate access to commercially valuable data <u>not</u> have resulted from industry influence in some form? The answers to these questions come in two parts. First, we will hear views of the project's origin directly from principals inside NOAA and review public documentation associated with the effort as they publish two formal Request(s) for Information from the vendor community (USDOC, 2014a; USDOC, 2014c) to obtain feedback on their ideas. Second, we follow the process as they evaluate this feedback inside NOAA, learning more about how they arrived at the final "concept of operations" and which companies to involve in that model, including observations by representatives of those companies about the process and their own interests, motivations, and challenges at the time. Together, we watch as NOAA moves from responding to requests for access from the commercial sector in Chapter 3 to attempting to shape the behavior of the private sector through their approach to distributing data.

Context: Institutional Focus on Secondary Use of Data to Stimulate the Economy

In 2013, there was a public focus by the Obama administration on the secondary use of

Open Data to a road to innovation and job growth (EOP, 2013a; EOP, 2013b). NOAA sits within the U.S. Department of Commerce and by that fall, there were significant Open Data initiatives in formation at the agency. As part of the Department of Commerce "Open for Business Agenda," Secretary of Commerce Penny Pritzker proclaimed the Department of Commerce "America's Data Agency" and rolled out a plan to emphasize use of Open Data as fuel for new startups, innovation, and job creation (USDOC, 2013). This agenda included reference to statistical agency data and also the BDP under a theme labeled "Unleash NOAA data through partnership" that said, in part: "We will launch a public-private initiative to unlock the economic potential of all NOAA data, with the goal of maximizing their economic and social potential, incubating new industries around climate and oceans information, and improving services and products available in the weather industry" (USDOC, 2013). How did things get here?

Filtering Down: Open Data inside NOAA

While NOAA already had a long history of sharing data with external parties, it is within the context of this Open Data initiative that one of the principals in the NOAA Chief Information Officer's office explains the origin of the project:

> We were essentially trying to spin up this idea as part of the Department's overall strategic planning efforts. They were looking for bold ideas in terms of getting the Department's data out and into the hands of people, liberating the data. (PRI-15)

The idea of creating jobs is a recurrent theme in these discussions and was a key part of the rhetoric of the larger Open Government initiative. As another stated, "[T]his was a project really, to help create American business, to create new opportunities and spur the economy and create jobs" (PRI-16). The NOAA Chief Information Officer (CIO) at the time, credited with starting the BDP project, recalled:

> Our promise in making the data easier to access and explore and use is that we could increase the value of the data, allow for greater insights, promote American innovation, and spur economic growth. Remember, at the time that we started on this path, I mean, job creation was huge and this could create additional industry. You couldn't do that with our existing infrastructure and funding…We knew that if we could position the data in a commercial cloud, we could remove that infrastructure bottleneck and allow value-added products to be created. (PRI-17)

He continued, explaining that he then pitched that concept to the NOAA Administrator and Department of Commerce leadership and they bought in (PRI-17). As the BDP project lead at that time explained, it was a top-down effort, with "[T]he NOAA Administrator at the time saying "Yeah, go do this" and some of her political appointees in the headquarters group actually worked with us" (PRI-18). They then decided to "[P]ull together a Request for Information to industry to try to figure out what we could glean based on the idea that we had" (PRI-15).

The Twist: Cost Avoidance

Admittedly, this view of the project origin is provided in hindsight and thus risks the danger of either "retrospective goal setting" or self-justification flavoring the recollections. Though it is generally supported by the principles and provisions included in the second Request for Information (RFI) that culminates the "origin" phase, as one might expect, the stories and rationale included in early presentations about the project do not all match perfectly, and there are several others that provide additional context that will turn out to be important later in this research. For example, one key NOAA Executive involved in the origin of the project recalled the catalyst for the idea of structuring the BDP so that NOAA did not pay for any of the work – a key to the benefits NOAA intended to derive from the project. It was prompted by consideration of a past offer made to NOAA by Climate Corporation, a company owned by Monsanto that makes use of weather data to provide precision forecasting for agricultural industries. He said that the offer was made in partnership with Computer Sciences Corporation (CSC), a large computer consulting firm that already did business with NOAA and it came before the first RFI had been developed. They proposed that NOAA provide NEXRAD (Next Generation Weather Radar) data to CSC and pay them to host it on their servers, so that they could make it available to Climate Corporation , who "[M]entioned various products from it that they could then use to grow weather (and climate) industry further" (PRI-19).

A high-level NOAA executive, one of three people on the original BDP team, recounted the experience [*emphasis added throughout*]):

> The Climate Corporation had a profit motive to build value-added products on government data. So they come to us with this idea. And, the idea ... okay ... and what we transformed it into is *something that they didn't suggest*, but something we'd been thinking about in parallel, without having it crystallize into a project. *And, the way it got twisted was this*: They were talking about this and we were getting all excited about the prospect of using cloud technology, right? You could put the data in the cloud and you could put the data next to the compute (cloud-based computing processing). And then it would make it much easier for private industry to use it. And the colocation of data and processing in the cloud becomes a force multiplier for the use of this information for injecting activity into the American economy….
>
> The dilemma? If you view that as "but we paid for this and isn't that a great thing?" that's problematic because there isn't any money. So our response to this idea, that if you just pay us to put the data in the cloud and you (they) get all this money - we twisted a little bit. And I said (to the others involved in the project), "*I think it's a great idea but I think we need to change it a little bit, I think they need to pay us* …Climate Corporation makes money on the products, they pay CSC to put up the infrastructure and the telecommunications to get our data out easier *so that our budget is held harmless*. If they're making money on it, our small costs for telecommunications and for the initial burden on infrastructure, especially if putting it in the cloud should be minimal and should easily be covered by the revenue stream from the folks who are making money on our data. *And when you then change that, now you should recognize that as the project we have today*. (PRI-19)

NOAA's Incentive: Avoiding the Increasing Cost of Increasing Government Quantification of the Natural Environment

Cost avoidance is, of course, a form of financial benefit to government that may also become vital to accomplishing parts of its mission, creating dependence on the source of the benefit – industry. While the anecdote directly above explains the origin of NOAA's attempt to shift the cost of distribution to industry as part of the BDP financial model, the potential strength of dependence here becomes clear when one sees the magnitude of the challenge NOAA faces in storing and distributing the increasing volume of the environmental data it collects. The volume of data being produced by satellites and other environmental sensors is increasing at an increasing rate, driving increased costs to NOAA for both storage and for distribution due to increasing demand for access to it by external parties. Something to remember, however: The needed storage space would have to be purchased from the private sector.

About eight months into the project, an executive at one of the collaborators made an informal assessment of the situation, reflected in a presentation made with the other principals at a national meeting of the American Meteorological Society in January 2016 (Haselden, 2016). She later expanded on her view:

> [T]he amount of data that's flowing, and the number of satellites that are up there, and then other sources, other sensors are already exponentially increasing. [I]n today's world, a very conservative estimate, which again, conservatively, does not include a lot of stuff that we are going to have access to even next year…the 2030 estimate was it would cost an organization about three million dollars a day just to contain, just to hold it. Not even to do anything with it. So, there needs to be a new way. There needs to be new technology or a new way to manage this, and include some of that in new policies, new processors to manage the data, because it's not going to decrease. (PRI-20)

At $3 million a day, this would grow to reflect $1.1B a year of what is, in 2018, a $5.9B total annual budget for NOAA (AIP, 2018). While early presentations during the run up to implementation of the BDP do not appear to emphasize this aspect of the situation, once the project was underway, few omit variations on the slides that follow, showing the volume of data and both increasing needs for storage at the archive and demand for access, in addition to about 200 petabytes of "model and other data" (Kearns, 2018). The rationale of cost avoidance in successfully addressing future access needs is also explicitly confirmed NOAA's response to vendor questions from a later project RFI discussed below:

NOAA is experiencing increases both in the volume and velocity of the data NOAA creates and in the number and variety of users of NOAA data -- developing an approach under BDP that will scale up to both increases in data produced and in applications by the general community is a fundamental motivation for BDP. (USDOC, 2014h)

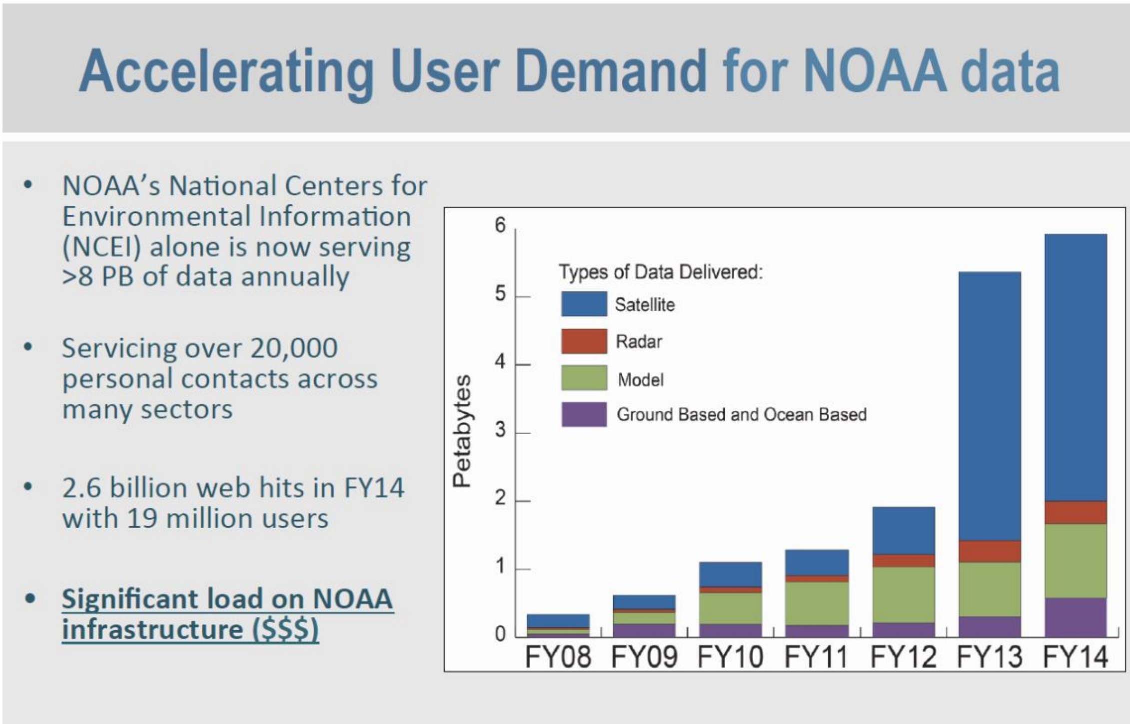


Figure 7 - Archival Storage Projections for NOAA Data (Kearns, 2016)

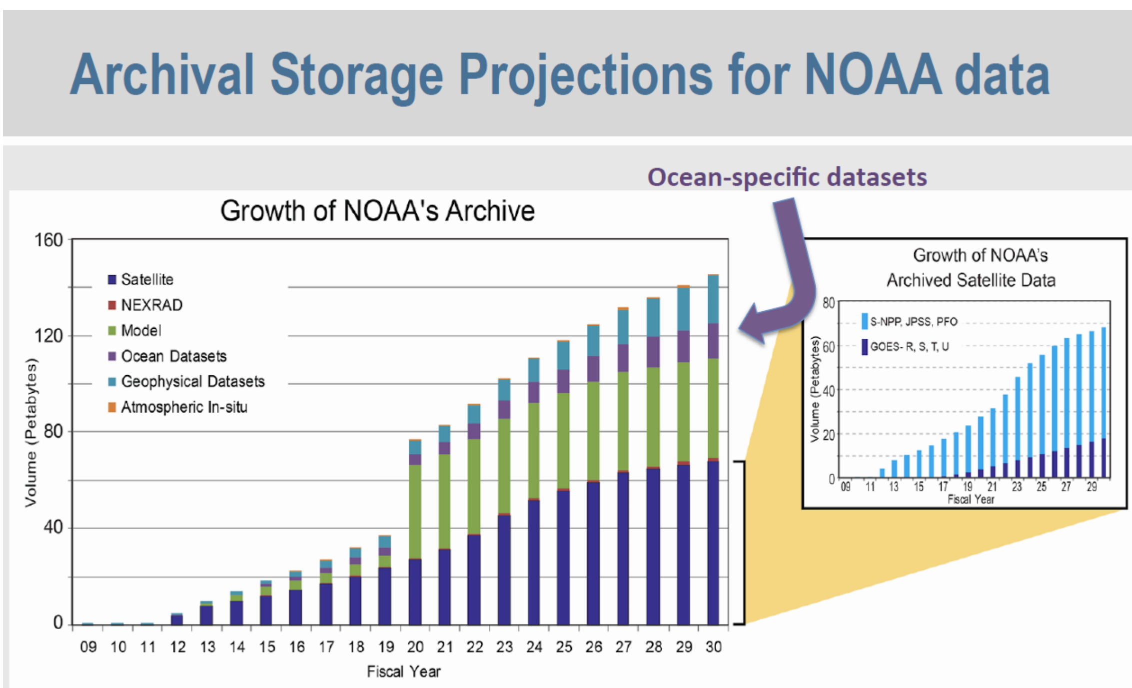


Figure 8 - Accelerating User Demand for NOAA Data (Kearns, 2016)

These stories advance our understanding of the forces of influence and the potential for dependence in the origin of the project in four ways. First, it demonstrates that NOAA resisted vendor pressure to grant access to NWP model data sought in the Reston Project and did not incorporate those requests in the BDP. Yet, while NOAA justified this based on equal-access policy, competing vendors apparently influenced NOAA to leverage that policy to stop an effort that they had allowed to get as far as having public briefings on its status where NOAA was represented and starting to develop a CRADA. This reinforces an observation in Chapter 3 about what I call "Boundaries of Convenience" set in policy through vendor influence, left vague enough that they are malleable and subject to vendor negotiation and pressure. Second, they support the idea that NOAA developed the project based on both its history of partnerships with industry and institutional initiatives at the Department of Commerce focused on Open Data, with a significant focus on job creation and economic innovation. So, there is an element of path dependence here, a collision between the historical mission of the Department of Commerce to promote industry, new open data policies claiming to have that benefit, and a NOAA history of previous collaborations with industry. Yet, third, it also highlights the discretion given to NOAA (discretion that appears similar to the latitude available to state government in the cases in Chapter 2) in choosing the specifics of distribution choices and arrangements – within the constraints of equal access - and that those can be in service of policy goals beyond just transparency and responsiveness to the public or commercial demands. Finally, presentations about the project show the impending pressure NOAA faces in the growing cost of data storage and distribution, apparently shaping their attempt to construct a partnership model that would offload distribution costs to industry users while continuing to fulfill or even expand their mission. This purposeful approach to distribution could, it appears – perhaps without fully thinking through the consequences - lead to significant dependence on the commercial marketplace for this subsidization over time.

From Mission to Market: Industry Feedback and the Design of the NOAA Big Data Project

So, we have seen and heard about the origin of the project from some of the principals involved at NOAA. Yet, once the idea took shape, they sought input – that is, opened a channel for influence - from industry and other interested parties to obtain their feedback in the form of a public Request for Information.

The First RFI

The internal discussions recounted above resulted in a formal Request for Information (RFI) published in the Federal Register on February 21, 2014 (USDOC, 2014a), with responses, due, after an extension, on March 31. While another, more detailed, RFI was issued six months later, the initial document helps frame the effort and demonstrate how the ideas discussed above were refined when put to paper for industry. For example, the project's purpose is summarized as "[I]ntelligently positioning NOAA's vast data holdings in the cloud, to be co-located with easy and affordable access to computing, storage, and advanced analytical capabilities….removing government infrastructure as a bottleneck to the pace of American innovation and enabling new value-added services and unimaginable integration into our daily lives" (USDOC, 2014a). The hyperbole ("unimaginable"), all part of the images in play here, is instructive of the environment and expectations in which the project was positioned. The RFI again references the project's goal of spurring economic growth, along with "tap(ping) the full potential of its data" (USDOC, 2014a). In the key difference from standard government procurements, it also establishes that any partnerships would be at no-cost to NOAA (the implementation of the "twist" previously discussed). The RFI was "for information purposes only, (is) not a request for proposals, and does not obligate the government in any way" and indicates the information submitted was confidential (USDOC, 2014a). While this confidentiality provision prevents analysis that could significantly improve our understanding of attempted influence – like analyzing the formal vendor feedback that was submitted - interviews with members of the project team involved in the evaluation that are discussed later in this chapter provide some insight into the general nature of the proposals.

The document contained seven sections: *Value Proposition, Data Infrastructure, Data Services, Data Management, Data Security, Partnership Methods,* and *Capabilities*. Among the questions included, NOAA asked industry to identify "high-value datasets," along with "the methodology you might use to determine the price paid by users for value-added services and products." Note that we see here, in the very beginning, NOAA's expectation that the private sector will have a good idea of the value of their data. Questions on data security, management, and infrastructure were fairly straightforward, predicated on concerns about maintaining the integrity of the data and effective delivery, and the *Capabilities* section asked questions to better understand the services the respondent could offer and their experience. At the time, NOAA indicated they planned to put out a traditional Request for Proposal (RFP) and asked for recommendations on "[W]whether NOAA should engage a systems integrator or multiple cloud vendors to accomplish this objective" (USDOC, 2014a). On its face, the table is left open for any and all combinations of approaches to be suggested, with the implication that the best thinking – notably, within the constraints of the model - would be proposed by the vendor community.

The Second RFI and BDP Industry Day

After six months of deliberation and "consultation with industry," the details of which remain mostly opaque due to the confidentiality restrictions of the RFI process, NOAA published a notice on September 26, 2014 (USDOC, 2014c) of an “Industry Day” for the Big Data Initiative to be held on October 17, 2014. It was immediately followed on October 3 with the contents of a new (second) RFI (USDOC, 2014d) that included a new 8-page draft of a “Big Data Partnership Statement of Objectives (SOO)” (USDOC, 2014e) and a spreadsheet for comments and suggestions. The Industry Day was attended by over 220 individuals representing 160 companies, along with dozens of representatives from NOAA and other government and non-government organizations (USDOC, 2014f). A NOAA presentation provided background on the principles and conceptual model for the project. Feedback on the first RFI was characterized as an “excellent response” showing the project concept was viable, and noting that “A majority of respondents provided examples with teaming/consortium arrangements involving multiple companies with various capabilities” (USDOC, 2014f). There was a “strong signal from industry” not to delay and to start incrementally, where a “critical mass of customers” along with “high-value datasets” already made the model viable. A list of high-value datasets identified by industry was also presented in the following slide:

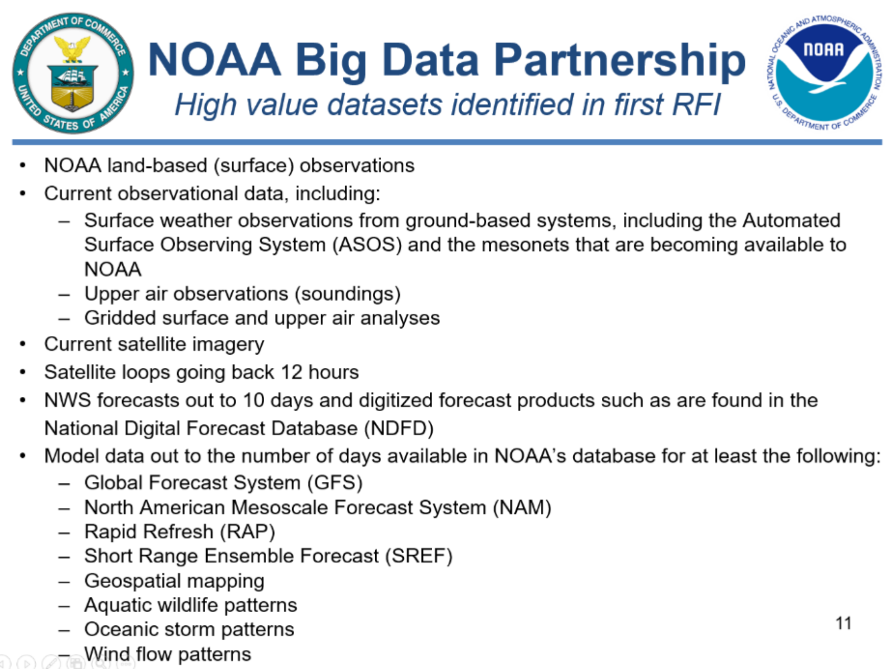


Figure 9 - NOAA BDP High Value Datasets Identified in First RFI (USDOC, 2014f)

The narrow focus on primarily weather-related datasets in this list is notable, especially as we look in the next chapter at the results of the project. For example, model data is emphasized, which we will see is not initially included in the implementation phase of the project, whereas Next Generation Radar (NEXRAD) – the first dataset that turns out to be deployed during implementation – is not mentioned. Also, as we shall see in a moment in some of the questions and answers from Industry Day, both parties still appear to struggle with what the most valuable datasets might be.

The Concept of Operations was also reflected in a diagram (Figure 10) of the roles of the parties, including policy and design principles. One "Partnership Principle" seems to have a direct lineage to the Reston Project: "No privileged private access to public datasets." And, noted with some irony given the confidentiality invoked in eventual CRADA described later in this chapter, is the principle requiring "Transparent management and operations of Open Government Data Service."

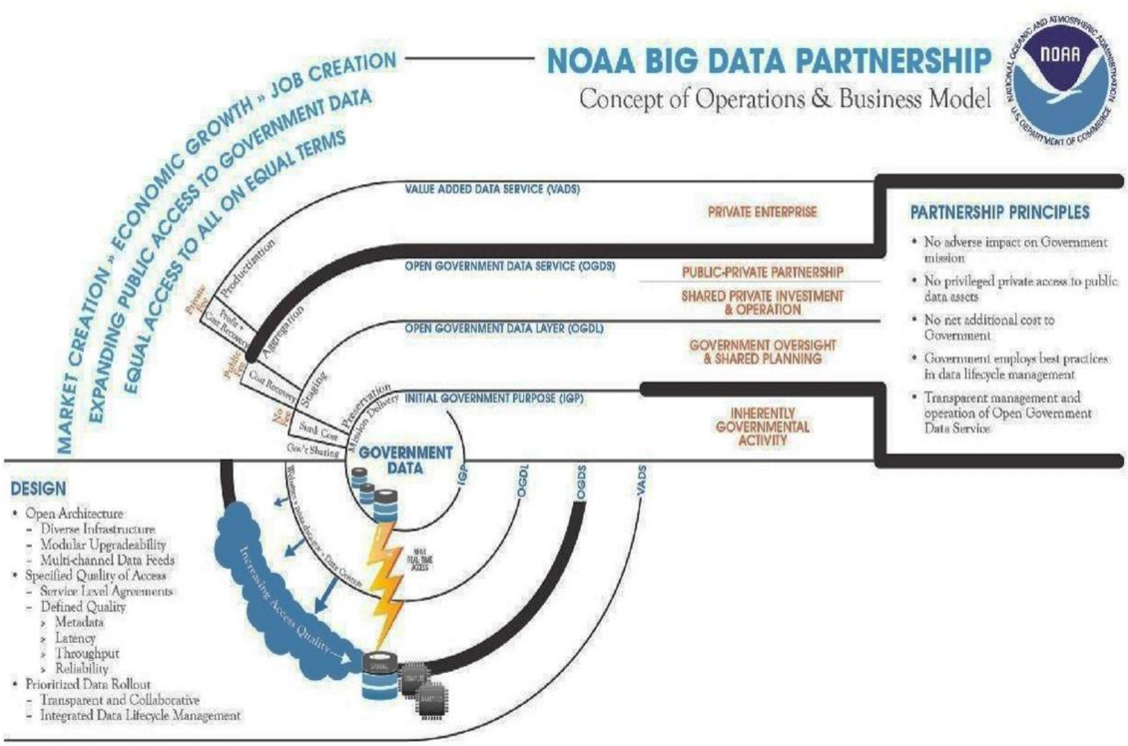


Figure 10 - NOAA Big Data Partnership: Concept of Operations & Business Model (USDOC, 2014f)

Finally, the presentation showed that NOAA expected a RFP to be issued by early 2015, and concluded with an overview of the new Statement of Objectives (SOO) document (USDOC, 2014e).

The BDP Statement of Objectives: NOAA's Interests Intersect with the Market

A close reading of the Big Data Partnership Statement of Objectives (USDOC, 2014e) shows it provides more detail than the first RFI, but in generally the same areas. Overall, it gives the impression that the requirements had been thought through in more detail, resulting in the addition of specific, stronger controls, such as statistical sampling to ensure data integrity (USDOC, 2014e), or, from an equity perspective, new provisions were included (2.1.6) that state "In order to enable the private sector to build business and create new jobs*, the Offeror shall not constrain the number of value-added service providers that can produce and provide products based upon NOAA data"* [emphasis added] and (2.2.4.1) "Partners may provide additional levels of access to NOAA's public data assets, but if they choose to do so the Partner must provide equal access on equal terms" (USDOC, 2014e). While strengthening the guidelines for the model, NOAA in turn placed a heavy emphasis on the limits of their commitment: "2.5.3 The Government will *not* provide a Service Level Agreement (SLA)" [emphasis added] continuing "2.5.3.1 This includes but is not limited to: access to metadata, defined latency, guaranteed throughput, reliability, assurances of continued access to a provided data, or a guaranteed response to an inquiry." This culminates in "2.5.6 The Government reserves the right to cease generating data and providing specific data sets without notification" (USDOC, 2014e).
A NOAA executive involved in the project explained the origin of this more determinative approach. Based on their experience, they already knew:

> (There were other) things we need(ed) to worry about. [T]hey all stem from the policies, like equitable access, latency – how important it is that everybody gets the data at the

> same time, and how we work with people in an equitable manner. And there's also the business question of them being able to distribute it in that manner, where we give it to them and they make it available for free, or at least at cost of free distribution of it. It was a business and a technical model. When we started ripping into the RFIs, it was like, "Holy cow, there's a lot of constraints around this Open Data policy. (PRI-15)

It was at this point, he described, that they began to think about the “rule sets” (PRI-15) they would need to include in the model, for example, data integrity and attribution (NOAA as the authoritative source of the data), security and performance impact on NOAA systems, equity in the timing of distribution – but also equal access on equal terms. “So, those are the things that came out and the nature of how it transformed a bit to be a pretty substantial thought piece around policy” (PRI-15). It is these rule sets that appeared in the second RFI and, later, the CRADA.

Reshaping the Market

Why are these provisions notable? To the degree that industry was running the show, one would expect considerably more deference to them in the latitude of operation allowed as well as at least some guarantee on NOAA’s part of the availability and reliability of the supply of information. In fact, in the Concept of Operations slide shown above (Figure 10), listed among the design principles were service level agreements and “defined quality” that lists the very subcomponents excluded from guarantees by the provision above: metadata, latency, throughput, and reliability. What is also important to note here, later echoed in the eventual CRADA, is the length to which these provisions show NOAA extending its own policy goals and interests as constraints (rule sets) on market design and operation. They can’t hand over their duties to a third-party, of course, without requiring them to enforce the same rules. But remember also that this is data the vendors can already download for free and do with as they please. While the development of the project still has many twists and turns ahead, these provisions are specifically referenced in later vendor interviews and lay groundwork that will help explain the project's results.

Industry Reaction to the RFI: Questioning Feasibility

While the responses to the second RFI are confidential, a spreadsheet was published shortly after the Industry Day with answers to over 100 questions posed by industry that provide further clues to commercial interests and concerns - and NOAA's reaction to them (USDOC, 2014h). While it likely falls far short of being fully representative of all the concerns and interests of the 160 firms that attended the Industry Day, they show that, over eight months after the project was initially announced, both industry and NOAA still seemed to be trying to understand how the model NOAA has put forth can be made to work and the number, nature, and value of NOAA datasets that would underpin its financial feasibility. As we seek to understand how vendor feedback during this process shaped the final form of the project and its operation, a few statements and exchanges stand out. First, there were multiple questions submitted that, in different forms, question the feasibility of the model. Examples include "Will NOAA compete with the provider, meaning will NOAA offer their data through other channels internal or external to the government?" "What benefits will the winning vendor receive versus obtaining public data through normal channels and creating their own value-added services? It seems like the winning vendor will be at risk long-term." And, "Why BDP if I can already access? Under open data policy M-13-13, NOAA is mandated to provide access to data and enrich that line. If so, I can access that data and do what I want with it. So why are you seeking partnership?" (USDOC, 2014h). (Note: M-13-13 is the federal open data policy (EOP, 2013b))

NOAA's Response to BDP Feasibility Questions

While NOAA's response differs slightly across these questions, they essentially say that they will continue to provide the services and data they already do – pointing out that they couldn't withdraw those without going through a significant process and evaluating that given their commitment to existing users (reflected in the requirements of their own partnership policy outlined in Chapter 3) – but will not create any new products to compete with offerings coming out of the BDP (USDOC, 2014h). NOAA's response to "Is there any other place where this proposed model has been stood up?" reveals that they have in mind an historical parallel to the project in a data distribution model for NEXRAD weather radar data. They state that, while they don't do anything exactly like this, nor are aware any federal agencies that do, "NOAA does operate legacy systems with some features which may be applicable to the BDP. For example, the National Weather Service established a real-time level 2 radar data service which incorporated some of the policy features contemplated for BDP, although it pre-dates cloud computing…" (USDOC, 2014h). In an interesting - and perhaps telling – act, this answer goes on to include a link to a master's thesis that traced the financial and policy components of that project, connecting it to the long history of public-private interaction at NOAA. Further, "real-time level 2 radar data" refers to Next Generation Radar (NEXRAD) data, which, as previously mentioned, turns out to be the first data type deployed in the BDP. When one casts the BDP as an evolution of distribution methods for NEXRAD data, it provides additional context for understanding mutual dependence and influence in terms of both path dependence and technological change. This and other aspects of NEXRAD data is the subject of Chapter 6.

Industry Interest in the Value of NOAA Data

The other relevant area consists of numerous questions about the data that NOAA has, its users, and its value. From the phrasing, there appears to be some expectation that NOAA will identify the relevant "prize" datasets, prioritize them and convey information about their value and likely uses/users to the respondents so that they can appropriately craft their responses. This is best typified by questions like the following: “What is the volume, variety, and velocity of the data?” “Will NOAA release estimates of the size of the data per collection time period for each type of dataset so offerors can plan appropriately before the RFP?” “Do you plan on providing a list of datasets prior to RFP release in a prioritized fashion, even if it’s non-binding or non-committal?” “Will all the data be available?” “What data volumes are we expected to handle?” And, most explicitly “Does NOAA expect to identify the initial high-value datasets prior to RFP release? These datasets will have a potentially large impact on teaming decisions among offerors” (USDOC, 2014h).

NOAA’s Vision for their Role in Identifying Commercial Value

NOAA’s response to these questions may have been surprising to industry, given that these industry questions, phrased as statements, can be taken to indicate that commercial parties are asking for this information to help make the project more successful and - one might assume - because at least some of them *really don’t know*. NOAA agreed to provide sizing information for at least “some key datasets as part of the RFP,” but their expectations about their role are best summed up in the following response:

> NOAA is contemplating that the proposed solutions(s) will advise a collaborative approach between NOAA and the data community in identifying the initial datasets. NOAA anticipates that designing and implementing a sustainable operational process of selecting data assets for release will be part of the evolution of the BDP. (USDOC, 2014h)

So, from the beginning, we see evidence of a disconnect between expectations of industry about NOAA’s knowledge of their own data and its commercial value and NOAA’s expectation that industry either already knew or would help them understand that. Perhaps the questions were submitted only by relatively unsophisticated vendors, but this lack of knowledge of available data and its value by government <u>and</u> industry is an important theme in our examination of how influence is (or is not) successfully exercised and the development of dependence via the project.

Shaping like a State: Policy Constraints, Market Incentives, and Managerial Discretion in the Design of the Big Data Project

When the NOAA BDP lead at the time received responses to the RFI from about 70 companies (PRI-15), he enlisted a larger group of NOAA managers to help in the evaluation and distributed the information across them for review. In addition, as part of the RFI preparation, NOAA used two loaned private sector executives through an existing program to help with the project. The participation of these individuals, one with experience in the weather industry becomes an integral part of the story that follows.

Responsiveness to Industry or Reengineering the Market? Part 1: Making Sense of the RFI Responses

From the beginning, the team had questions about the operating rules to put in place between NOAA and private sector partners for the project. As Executive 1, one of the loaned executives recalled, it started…

> [A]t the very end of Industry Day…the last question in the question and answer session, someone in the back raised their hand and said, “So if this is an RFP, if this is a contract, what is it exactly that you're awarding? And everyone said, “Uh ...can we get back to you on that?” And this I think is really where I started to get involved in it is in that process of brainstorming, “How is this going to work? What are we going to do? What are we giving the person or the company that gets this contract? What do they actually get out of this?” (PRI-21)

The companies responding to the RFI fell roughly into three buckets: A large number of “standard” government contractors” (the federal-focused “Beltway Bandits”) (PRI-21); A few infrastructure-as-a-service (IAAS) providers (firms specializing in hosting data in the cloud), although “there wasn’t a lot of interest there” (PRI-21); and, large industry players in the weather and climate industry. These industry groups were considered the front runners at first (PRI-21). The content of the responses reinforced NOAA’s picture of the existing commercial ecosystem for National Weather Service data that supported the multi- billion-dollar weather industry. There were aggregators that had expertise in moving data to a place they could re-distribute it from; Platform providers who hosted it – Infrastructure-as-a-Service (IAAS); Companies who added value to the NOAA data to generate a “second order” product; and then commercial customers of that second order product (PRI-15).

While, as noted above, there were few responses to the RFI from IAAS providers, the executive running the project from the NOAA side explained how the team came to focus on them as key to the eventual model. In reviewing the proposals, they found that some companies concentrated just on one of the elements of the ecosystem described above, but many also looked at the “full value chain” , which led the vendors to propose teaming up with other firms (PRI-15). One commonality was that those teaming up all referenced a single cloud (IAAS) provider hosting the data for the partnership. In his words:

> The interesting part was you had certain companies that would suggest using Amazon as the (cloud provider) infrastructure, or others that would suggest using Microsoft. They (the RFI responses) had all the common components, but at the end of the day, when you analyze the responses they all pointed back to these Infrastructure-as-a-Service providers. If you look at the distribution of them, and you looked at the time as far as how large those infrastructures and service providers were in terms of market dominance, it was pretty consistent across the RFI in terms of how the majority, what number of responses referred to one provider versus another. (PRI-15)

Here, then, you can see the decision makers considering industry responses and the structure of the information market in which NOAA is participating. The explanation of their thought process as we move forward illustrates how industry proposals are considered, along with the interaction of policy and administrative discretion in their decisions. *At this point, the impact of industry influence seems to be felt only in the proposals from which NOAA officials abstract their own* *view of what's possible*.

Responsiveness to Industry or Reengineering the Market? Part 2: Disrupting the Market

In considering the idea of distributing data via the cloud hosting platforms (IAAS) as the core of the model, the analysis team identified another decision to be made - whether to focus the agreements with commercial partners on specific domains of environmental data or across all of them. The NOAA executive running the project continues the story:

> So we looked at – how do we stack this thing, do we stack our "ask" in more of a domain-specific kind of way. In other words, if you had a full end-to-end chain, would we try to incentivize folks to bite off the weather piece? Award to a group of people and say, "Okay, you go work off of the weather service, and you go work off of fish data, and you go work off of satellite data"? You get communities of interest around those areas.
>
> We thought about that. We're like, okay, so you have these typical verticals like major weather companies that already have groups and consortiums around the sharing of weather data. What was really disruptive was, what I think was disruptive, is when we really thought about it and said, "You know what? The power of this is to make it really extensible and think ahead to all of government. Why not stack it in such a way where the purpose of this, the real power of it is if we can replicate the business model and feed multidisciplinary data into the same place, such that people can do analytics off of it in the same place?" (PRI-15) (See also the concept of "Disruptive Innovation" in business popularized by Bower and Christensen (1995) - the implied connotation for his use of this term).

Expanding further on the concept, he explained:

> You're bringing together computing and the locality of the data with the computing, and then you have these data pools where now you can combine fish data with weather data, with satellite data, with transportation data, with health data. That in the long term, it didn't make a whole lot of sense if we bored out these little domains and start these little fiefdoms of market share or (replicated) that self-sustaining market.
>
> We focused it and said, "Well, it seems like the power of it is to focus it around getting these pools of data." When you look back to it, you say who are the people that can help us do that? That's where we landed back on these Infrastructure-as-a-Service providers, like the Amazon, IBM, Microsoft, the Open Cloud Consortium. They had more of an interest in making Open Data available on their systems and having the associated infrastructure around the data to create the market space versus incentivizing (someone) like a Weather Company. Why would they want to pull in health data? Maybe for their purposes, but for everybody as a whole? We started thinking about it from that perspective. *It became really interesting because now all of the sudden people have to*

> *think about things in a different way where you typically had these redistribution points for weather industry or for satellite or other locations that now all of the sudden have to play in a different way in a community* [emphasis added]." (PRI-15)

Others central to the decision agreed that this was the rationale. One of the loaned executives helping with the project confirmed:

> So the idea, obviously, is that infrastructure-as-a-service providers, their role in all of this, the way that they make money is if lots of people use the data. Their incentive is very different from the incentive of a Weather Company. (PRI-21)

Another key NOAA executive involved in the decision added:

> Also, around that same time we realized - well we thought we realized - that the right part of the market to deal with was not the resellers themselves of the data, or like the people were actually going to be using the data, but rather the platform intermediaries (IAAS providers) because they could take all of our data, not just some of our data. It wouldn't be cherry picking. They could take it all and then support the people who do the cherry picking. (PRI-19)

From these first-person explanations, we see that - rather than influence coming from the private sector - government officials are using their understanding of the information market and private sector incentives and logics (accurate or not) - within existing policy constraints - to construct a distribution model that they believe will further their mission. In this case, the underlying assumption appears to be that NOAA has the leverage to do this because they have the valuable resources that industry either already depends on or, optimally, will come to do so. An important aspect of this "experiment" is also that - while it may portend major changes in the future – there is no commitment to stop any of the existing methods by which commercial firms presently obtained the data, a point made explicit in NOAA's answers to RFI questions. From this, one might take two things: First, successful or not, the BDP is unlikely to disrupt current operations of any current user of their data, and, second, this same fact seem to make dependence on distribution via the BDP an option or opportunity for a vendor, versus a threat to their bottom line. When coupled with NOAA's stated "best effort" stance in committing to data quality and delivery – versus a guarantee - the incentives for industry participation seem less than compelling.

Responsiveness to Industry or Reengineering the Market? Part 3: The Data Alliance

To accommodate the idea of the communities of interest that existed around different domains of data now subject to the constraints of the IAAS-centric delivery model, NOAA developed the concept of "data alliances." The concept in its refined form is explained on the NOAA Big Data Project website (http://www.noaa.gov/big-data-project-how-to-participate) under a section titled "How to participate":

> Each data alliance is anchored by a participating Infrastructure-as-a-Service (IaaS) institution (a "Collaborator"), and represents a market ecosystem built around one or more of NOAA's datasets. The data alliance is a flexible concept, and may consist of larger companies that represent various economic sectors, such as the weather or insurance industries, specialized small business, value-added resellers, entrepreneurs, researchers and non-profits, etc. The data alliance structure allows market forces to act on the identification, extraction, and development of NOAA public data resources, and provides a mechanism for interested parties to work together to develop new business and research opportunities. The organizations comprising the ecosystem are free to participate in multiple data alliances. (NOAA, n.d.)

On its face, then, the concept of data alliances allows everyone to participate. And, as conceived by NOAA, there is an assumption that all the platforms will have an incentive to form or identify ecosystems of users (aka Data Alliances) to achieve a financial return on their investment in providing both hosting and access for free. As Executive 1 put it "(The) idea was to target the infrastructure-as-a-service providers as being core anchor partners within this data alliance" (PRI-21). Fostering these communities was also a requirement addressed in the eventual agreement – the CRADA – made with the IAAS providers. That said, in an earlier conversation, we see NOAA's awareness that such ecosystems and alliances already exist, and the BDP couched as a way to encourage cross-disciplinary / domain efforts that could invent / reinvents uses and markets via this "disruptive" (PRI-15) innovation in data distribution.

From RFP to CRADA

NOAA executives had laid out the rule sets, the responses to the second RFI had been received, and they had made the decision to make the IAAS providers central to the operation of the model. According a NOAA executive involved in leading the procurement, as the expectation had been set at Industry Day for an RFP to be released early in 2015, the question became “How in the heck are we going to award something like this?” (PRI-15) Executive 1 agreed that part of the challenge was the procurement vehicle: “How do we convince them to bid on this “quote unquote” RFP? How do we guarantee that what we end up with is an IAAS provider if we're evaluating RFP responses and some IAAS providers don’t bid…?” (PRI-21 The executive asked if there were a way to approach the IAAS providers directly and someone suggested the Cooperative Research and Development Agreement (CRADA) (PRI-21). The rationale was that, while it was usually used for scientific research, this was a type of *business* research. Once that decision was made, it was “four months of lawyers,” along with outreach directly to the IAAS providers to complete the deal (PRI-21).

“What the Hell Are You Talking About?”: Selling the BDP Model

The move to focus on the IAAS providers via the CRADA vehicle was a significant watershed in the Big Data Project. While 70 vendors had responded with proposals, one can see through a recounting of the evaluation process above that NOAA – of its own volition – decided upon a “concept of operations” for the distribution of data that reflected their own view of how a market might work to achieve their goals – rather than pursuing any of the proposals made by industry. For better or for worse, however, this left them with the task of contacting the IAAS providers to explain and interest them in the proposal, vendors who, in some cases, did not appear (all responses are confidential) to have even responded to the RFI. Four IAAS providers were targeted based on their top ranking in a technology advisory service (Gartner) publication (PRI-21). The path ahead now lay in finishing the proposed CRADA and persuading these IAAS providers: Amazon, Google, IBM, and Microsoft, along with a fifth academic cloud provider who had responded to the RFI, the Open Commons Consortium (later renamed Open Cloud Consortium), to participate.

While NOAA saw the participation of multiple providers as central to success of the model, given the public visibility into the competition, Executive 1 explained there was another incentive to gain agreement: “It would have been bad optics, right? It would have looked bad to the American public if we had just one” (PRI-21). The varying constraints and considerations here, then, expanded beyond a strict bottom line evaluation and included the need for equitable access (or, at least, its appearance). And, while the project may have not been undertaken in response to pressure from industry, we will see that the attempts to persuade the IAAS providers to participate and the consequent expectation-setting are shaped by everything from their organizational structure to their business model to their understanding - or lack of it - of both NOAA data and the market for it.

What was NOAA up against? From a commercial point of view, the loaned private sector employee (Executive 2) that had a background in the weather industry, shared that observed that before he had joined the project:

> [T]he project as it was defined at that point was to do a partnership and do an RFP, a real government big RFP, but in this weird, no-cost-to-government way. That just seemed very complicated and made not a lot of sense. Who wants to sign a contract with the federal government to provide a bunch of stuff for free with all the onus of contracting? (PRI-22)

And now, of course, the rule set included in the second RFI had only expanded these requirements. Ultimately, NOAA was now dependent on the success of these IAAS providers and their enlistment of commercial customers for the success of the project, avoiding increased costs of data distribution while creating jobs and increasing exploitation of value latent in their data: All that would have to be accomplished by industry. As a practical matter, then, the IAAS provider's response to a de facto NOAA policy in the form of a CRADA set the parameters for how the project could work in practice, regardless of the origin of the idea.

Challenges: The Operating Models of Commercial Organizations

One of the first challenges encountered in convincing the IAAS providers to participate was that their organizational structure and posture toward government was not designed to evaluate and respond to a proposal that didn't involve sale of a service or product. In contrast, born of an effort to enable researchers to use collaborative cloud computing across academic institutions (Grossman, 2015), Executive 1 recalled that the Open Commons Consortium was "all over this" and viewed it as an opportunity (PRI-21). However, to engage the other four hosting companies, the NOAA team started by convening meetings with each company's federal sales representative who, according to Executive 1 had a little different response:

> They're like, "What the hell are you talking about?" For each of these companies, we had to kind of work our way through the sales people into some other group within the company. There was no really clear idea on either side what group that might be. It started with the sales people. The sales people said, "Uh, okay, let me call my boss." The boss hands it off to someone in marketing and marketing hands it off to someone in some other group. We got passed around a lot and had to have that conversation with each of them in which we sort of described the idea and talked about specific examples of existing data sets that were being used by various companies. (PRI-21)

The companies under discussion are some of the largest in the world. While NOAA had ambitions to usher in "disruptive" innovation (PRI-15), this episode begins to peel back the layers of assumptions and motivations that result in commercial markets "as they are" and pulls NOAA deeper into an attempt to reshape their operation and incentives to what it perceives as mutual advantage.

Other obstacles to engaging NOAA's proposal may not have been visible to them. One IAAS provider that eventually agreed to participate described the challenges from the inside:

> Since there was no revenue from the program - there was no payment either way - we had to look to our practitioners to help us with the different skill sets, and ask them basically to do it as additional work above and beyond their normal day-to-day work. They're also basically volunteering their time. Because we aren't making any money. They're not billable. What we ask them to do is they have to bill their full-time work to justify their positions, but then they are very interested in NOAA and weather and big data. There's things that intrigue them about the program, so in addition to their normal day-to-day work, they also supported the CRADA program. And that way, (we) aren't out any money – it was all above and beyond. (PRI-20)

Challenges: Selling the BDP to the NOAA Organization

Legal staff had determined that, instead of the head of NOAA signing off on the CRADA, each of the assistant administrators inside the organization had to do so. So, there was also a sales effort inside NOAA similar to the external partners:

> There were some who just said, sounds great, where do I sign? And then there were some where, again, the day before the deadline on this, we weren't entirely sure that they were going to sign the CRADA. I think the biggest fear internally was that the data would be misused. It was sort of the fear that, and I think it's a valid fear, that data can be manipulated or used ... a lens can be put on data to justify or quote unquote "prove whatever you want to prove."(PRI-21)

In the current process, the customer service staff at the National Center for Environmental Information (NCEI) would, for the most part, know who was downloading or using the data – and scientists could be put in touch with them, ideally, if there were questions, and explain the data source or its structure. In the future, if it were obtained from a cloud provider, NOAA would have no visibility into who the requestor was or their probable uses. This lack of information, combined with wider use, also makes mapping interests, dependence, and influence more difficult, complicating the ability to counteract potential problems with policy in advance.

Many Hands, Many Cards to Play: Entangled Influences, Motivations, and Incentives in Information Markets

It seems clear that the IAAS providers were not organized to quickly take advantage of the model NOAA proposed. They also lacked an organizational focus and expertise to easily identify NOAA data that would provide a sufficient return on investment to make the model work. Finally, looming over the discussions was the 800 lb. gorilla, frankly summarized by Executive 1who had moved to managing the project as they completed the procurement:

> One of the interesting things about that CRADA, and I don't know if anyone has put it to you quite this bluntly, is that there's absolutely no reason to sign a CRADA to do this. Any Infrastructure-as-a-Service provider could have done this if they wanted to. Any company could do this if they wanted to. It's public data. They don't need NOAA's approval to host it and make it available to other people. (PRI-21

While the IAAS providers were closely focused on the legal agreements and the cost of any commitments included in them, subsequent discussions with their representatives revealed that their interest – such as it was – came less from confidence in the proposition and more from the value of enhancing relationships with NOAA, both existing and prospective. As the lead for one of the IAAS providers that joined the CRADA saw it:

> I think the value of the CRADA (to our company) is not access to the data...it’s not that that made it an interesting program - it's the model of working directly with NOAA. Because when we ask about the fisheries data, when NOAA helps make a connection to somebody who needs fisheries data, those are connections and relationships and the other communication that the CRADA enabled that we wouldn't have had if we just put our request in the queue. (PRI-20)

An executive for another IAAS provider that also participates conveyed that they’d made a similar assessment at the time:

> I think a lot of people (internally) thought this would lead to business and I think ... I get the sense there are people on NOAA's side that thought that this would lead to contracts with the cloud providers, and I know certainly from our side there's sort of the assumption that NOAA wants to do this and if we play nice, then we will be favorable for

> future procurements and things like that. That was kind of the business rationale for entering the CRADA. (PRI-23)

On its face, it appears then that the scope of the information market as it was conceptualized by NOAA and the "whiteboard view" of IAAS providers as intermediaries - whose interests would be focused on the revenue opportunity for selling computing cycles on Open Data rather than the larger ecosystem of their dealings with government - was unrealistically narrow. For example, Federal expenditures for cloud services were recently estimated to increase from $4.9B in fiscal year 2017 to $6.5B in 2018, representing 8 years of increasing investment (Criste, 2018), business that four of the IAAS providers would have a direct interest in. Closer to home, another illustration of the diverse and dynamic interests of industry occurred when - about six months into the project (October 28, 2015) - IBM (a participant in the CRADA) purchased The Weather Company (Weather Company, 2015). This development is notable for a number of reasons. First, in an odd turn of events, the same Weather Company that was part of the Reston project that was shut down due to the "privileged access" that they had been accused of trying to establish had now been brought in as, essentially, part of one of the firms who were awarded an agreement that was not publicly bid via RFP. Rather than implying impropriety, this simply highlights that making assumptions in setting policy to ensure fairness /equity and then anticipating dependencies – and corresponding avenues and incentives for influence – is very challenging in an information market where the business ecosystem continually changes. These layers of motivations, incentives, and relationships are difficult for government decision makers or the public to anticipate. Witness this statement by a person employed by The Weather Company at the time:

> When the Amazon program got started, I was heavily involved with that and really gave them a whole lot of strategic direction on what we could do here. It wasn't long

> thereafter that IBM bought The Weather Company. That made it difficult for me to continue to be involved strategically with Amazon. (PRI-11)

When data becomes of commercial value, while it may be understood that concentric circles of use and interdependency are created far beyond the horizon visible to a government agency – in fact, in NOAA's case, expanding data use was one of the BDP's stated goals – the lack of visibility into this ecosystem means that initial assumptions may be difficult to validate or need to change in both the short and long run. This also complicates anticipation of where and how influence may be exercised based on changing alliances and incentives.

In a similar vein, the motivation behind a government project is not always clear to the vendors themselves. Consider this anecdote about a situation involving competition between at least two of the IAAS providers, IBM and Amazon (Microsoft also apparently was involved at one point as well) (Konkel, 2014). Prior to the BDP, there had been ongoing conflict over a large cloud computing procurement by the Central Intelligence Agency that involved lawsuits over the evaluation of proposals and resulted in Amazon winning a contract over IBM that was worth potentially $600 million. In light of what happened, one current representative from a collaborator related this story, passed on from a principal involved earlier in the project:

> One sort of theory that we've had of the motivation for the CRADA was that it was a chance for NOAA to do a very public vetting of all the infrastructure providers, so that there's no way that if they procured anything from one of them that one of the providers could then turn around say, "Hey, you weren't fair to us." I do think that ... I mean, regardless of whether or not that's all true, it's obvious that procurement laws that make it difficult for agencies to procure cloud are informing what's going on with the CRADA. Because basically, in one sense, what the CRADA boils down to is NOAA needs to be able to explore what it's like to use a public cloud environment. They just have no way of doing that on their own, so they have to go through this Byzantine legal agreement. What they should be doing is buying the cloud - It's quite cheap compared to what they do now - and experimenting and exploring, but it's very hard for organizations to do that. (PRI-23)

Far-fetched? The award of an unrelated contract with the Department of Defense for cloud computing and related services worth an estimated $10B went on hold in late 2018 due to litigation over the proposal to award it to a single vendor. Microsoft bid, Google backed out of the competition, and IBM and Oracle filed protests (Nix, 2018; Knapp & Corrin, 2018).

This dance of uncertainty, multiple motives, and one action that advances multiple agendas may be no different than other contracting relationships entered into by government. But, because in this case, government is attempting to be the supplier of goods - not the procurer – while attempting to restructure distribution of those resources in a pre-existing market using a new model - the operation of influence may be wrapped up in several agendas, not all related to the data itself, but to strategic positioning with regard to other "streams" of value available to business – or even government - several moves ahead of the current decision. While dependence exists, there is also interdependence of goals, markets, and organizational knowledge and capabilities, making influence on data and policy both difficult to unpack - and to exercise.

Competitive Incentives and the Known Unknowns

Eventually, the NOAA team began circulating the draft CRADA among the vendors so that they could agree on a common document. The former NOAA project manager is bound by the confidentiality provisions of the CRADA, but indicated that of the four companies, at that point one was going well and they had been going back and forth with the lawyers for several months. From NOAA's point of view, another was "interested, but not really getting down pragmatically to the nitty gritty of what would need to be done." And, the other two were "…kind of nowhere. We were really having to push them to even keep talking to us" (PRI-21). Questions about the data and its value also came up again – echoing some of the same questions noted earlier in the RFI responses. After the potential collaborators grew more serious and began to ask questions, the answers were slow in coming. The BDP project manager at the time recounts the situation:

> [A]t some point we did kind of poll industry (the kinds of companies eventually envisioned to be part of the data alliances) and say "What datasets would you like to see? What would be most interesting for you that you know we have?" This got us into a death spiral. So we said, to industry, “What are you interested in?” They said "What do you have?" We sent back, "Well - everything.” They said, “Well, we need to see what you have to know what's interesting.” Then we internally tried to come up with some sort of catalog of all of the data that NOAA has...As it turns out, it's impossible. It's in all sorts of different formats at all sorts of different levels of granularity. So, that catalog was kind of a mess. What we did have was at NOAA.gov.... That was a preexisting catalog that had been put together as part of the Obama Open Data policy. But again, and to be clear, this is nothing against the person who did this - they did a fantastic job with very low resources - but because there is so much data at NOAA, because it's so diverse - coming up with that list of “Here's the data we have” was a lousy deal.
>
> Again, it's a problem that I'm not sure could be solved without sort of a top-to-bottom grouping over the data standards and data policies. (So) the best we could do is say, "Well, here's some possibilities, are you interested in these? So we kind of kept going back and forth on that.” While the information was definitely being used to tempt the Big 4 (IAAS providers) into this, it was information we were trying to figure out anyway just to get a handle on what scope of what we were asking really was. (PRI-21)

Toward the end, the NOAA team leveraged the natural competition between the vendors to get them to sign. As the project manager (now Executive 1) put it “Certainly, I think that the second vendor signed because of the first, then the third due to those two…there's obviously all sorts of communication going on in back channels, but we were saying, we've got one partner ready to sign. So, if you guys want in, get in now” (PRI-21). Eventually, the CRADA was signed by all five parties, the last one within 12 hours of the final deadline. Using vendors who didn’t bid, with each uncertain of the other’s agenda, the value of the data – or even for sure what it was – and no guarantee of its availability, nor whether there was enough interest to fund the model, the project was ready to go.

Policy Experiment: The BDP Cooperative Research and Development Agreement (CRADA)

The BDP CRADA has been referenced a number of times in the discussion above, with only a brief outline of the final framework provided. However, the specifics of this agreement are important for at least two reasons. First, the terms are those under which the implementation phase of the project discussed in the next chapter must operate. Second, these terms, agreed to by all the Infrastructure-as-a-Service providers, extend government policy – albeit experimentally – in an attempt to influence the behavior of commercial (and academic) users of government data with the purpose of increasing their use of and dependence on it and, over time, avoiding or subsidizing NOAA's costs to make the data available. These terms, then, spell out a relationship intended by NOAA to foster mutual dependence on the commercial use of information – the subject of this study.

CRADAs are often used by NOAA for joint research projects with industry that may produce commercially valuable intellectual property to define the rules, relationships, and expectations among the parties, along with ownership and use of what is produced (NOAA, 2018). In this CRADA, NOAA's states its rationale for using this type of agreement as follows:

> (13.) NOAA has developed a desired framework for how it envisions such a public-private partnership might operate. After significant market research and interaction with industry, NOAA has determined there are many unspecified aspects of this framework which only can be effectively answered through collaborative research, development, and experimentation with industry.' (NOAA, 2015)

The following aspects of the agreement are directly relevant to the study of mutual dependence and influence:

Intellectual Property

In considering the constraints embodied in the CRADA, it should be noted that "CRADA data" – as it is defined in the agreement - is Open Data that has no restrictions on public access or use. And, the definition of what can be considered intellectual property – a "CRADA Invention" – is limited only to the inventions first reduced to practice in support of the project "by embodiment in data extraction technologies" - essentially processes to get data from inside NOAA to the collaborator platform (NOAA, 2015). One can imagine that being able to do this in a high-speed or highly efficient manner with high data integrity could be a competitive advantage. But, it also highlights the very narrow way the role of the platforms and their interests is envisioned. The project ground rules are conceived such that the collaborators are simply intermediaries – it is instead their customers who would "productize" the commercial value of this data via Data Alliances, with collaborator revenue coming primarily from customer use of their high power computing assets. As we saw previously, their interests can be much broader than this.

Proprietary Information

One provision of the CRADA protects "Proprietary Information," defined as:

> [C]onfidential scientific, business, or financial information, including data created under this agreement solely by the Collaborator at the Collaborator's research facilities, which may embody trade secrets provided by the Collaborator to NOAA in the course of this CRADA, and developed exclusively at private expense... (NOAA, 2015)

As a practical matter, when taking into account NOAA's dependence on the goodwill and continued interest and resources of the collaborators for the project's success and apparent deference to the broad nature of what might be considered "trade secrets" or proprietary information by commercial parties in this unusual area, NOAA and the collaborators have chosen to keep most of the details of the internal operation of the project confidential. This includes documentation and detailed discussion of meetings or related materials, making an empirically-based understanding of some aspects of the project and its operation very difficult. For example, in the implementation phase: knowledge of collaborator return on investment, an understanding of what data NOAA experts promote to the collaborators and how, which datasets have been rejected, as well as any data actively solicited by industry - and why. While there is heavy emphasis on "equal access on equal terms" to data (see below), this agreement grants five parties - to the exclusion of the rest of industry - privileged access to NOAA expertise in performing research and identifying opportunities to commercially exploit the value of public NOAA data, none of the results of which were shared with industry or the public. While some collaborators confessed above that their interest in the project was more about access to people than data, when cast in this light, it appears that this is really what the CRADA gives exclusive access to - *not the data, but to NOAA expertise to help identify its potential for commercial exploitation*. This claim receives further support as we examine the execution of the project in the next chapter and finally comes to be explicitly stated in yet another RFI released as the CRADA comes to an end.

Extending NOAA's Rules of the Game to Private Industry

An appendix to the CRADA addresses the "technical objectives" and requirements for the parties – many repeated verbatim from the Statement of Objectives in the second RFI. Elements key to mutual dependence and the exercise of influence include (all referenced below from the final CRADA (NOAA, 2015) unless otherwise noted):

- A provision (1.1) for NOAA to collaborate in identifying datasets to be hosted, creating a channel for influence by industry and the development of mutual dependence.
- The requirement for "Equal Access on Equal Terms" (13, I, 2), including a specific prohibition against the collaborator granting "privileged private access." The concern about privileged access hearkens back to the issues that derailed the Reston project. It again begs the question of what would motivate a collaborator to accept these constraints *for open data they could obtain freely at any time and distribute any way they wished without entering into any agreement with NOAA* – and reinforces the idea that it was less the business opportunity

here and more the potential for future business and access to people and expertise.

- The work is to result in "no new net cost to government." This is a key principle underpinning the model and was expanded upon in the Q&A associated with the second RFI:

> Our "no net cost to the government" requirement should be understood in this context - BDP may generate cost savings to the government which will offset the government's costs for establishing and supporting connections to winning vendor(s), thus creating an overall "no net cost to the government" outcome." (USDOC, 2014h)

This clearly defines cost avoidance as a benefit government will depend on from the relationship.

- A Collaborator cannot charge for the data and it must be "available to the public without restriction on use" (I.,1,1.3.2) – essentially extending both NOAA's and federal open data policies to define the actions of commercial entities who were parties to this agreement.
- The CRADA further states (I.,1, 1.4) that "In order to enable the private sector to build business and create new jobs, solutions should not constrain the number of value-added service providers that can produce and provide products based upon NOAA data." It is notable that NOAA extends the provisions of the agreement *to specify how the private sector Collaborators should distribute data in a way that creates jobs* – in contrast to the assumption that this result would be a natural outcome of granting them entrepreneurial flexibility. It is hard to imagine a more direct intervention by NOAA in an information market than this.
- The expectation that the collaborators will work with industry alliances made up of organizations <u>not</u> named as collaborators to "advance the outcomes from this CRADA with more robust and potentially diverse solutions." This, along with similar language, codifies the concept of Data Alliances discussed during the RFI evaluation, positioning these collaborator platforms as intermediaries to industry to distribute and otherwise facilitate making use of the datasets selected in the project. A key expectation by NOAA follows:

"This cooperation also enables NOAA's ability so support a greater number of participants" – of course, potentially defraying NOAA's cost to do so. Notable here is the artificiality of the construct. NOAA is requiring an IAAS provider to build a market - a set of customers - for individual datasets. In publicizing the value of its data, NOAA states it is already "Services 20,000 personal contacts across many sectors" (Kearns, 2016). They seem to be saying "We need you to rebuild that market in such a way that you attract a critical mass of dependence on processing this data on your platform so that you may produce enough revenue that we may avoid the costs of that distribution going forward."

- Again taken directly from the Statement of Objectives in the RFI, "4. NOAA will provide datasets on a best-effort basis" and "reserves the right to cease generating datasets without notification." This provision provoked questions about the feasibility of a business model based on such a weak guarantee during the RFI Q&A. As a practical matter, when coupled with the limited term of the CRADAs, it provides a strong disincentive for industry to invest significant time and resources in the BDP when they could instead maintain current supply lines for data where they have already achieved commercially sufficient reliability.

Taken as a whole, these provisions suggest that this mechanism is – in effect - being used to turn these IAAS providers into an extension of NOAA, intervening in the market with rules that mimic the same restrictions NOAA has (make data available to everyone at no cost), while directing them to expend additional resources to attract new market participants to their platform so that NOAA may avoid having to pay for the hosting of its data. What will happen next?

Award of the Big Data Project

In April 21, 2015, exactly fourteen months after the release of the first public Request for Information on the NOAA Big Data Partnership, a press release was issued to announce the

CRADA awardees. The following excerpts summarize that announcement:

> U.S. Secretary of Commerce Penny Pritzker today announced a big data project with Amazon Web Services, Google Cloud Platform, IBM, Microsoft Corp., and the Open Cloud Consortium to explore ways of bringing the Department closer to its goal of unleashing its vast resources of environmental data and delivering on one of the Department's key priorities – transforming Department data capabilities and supporting a data-driven economy…. These collaborations, established through Cooperative Research and Development Agreements, or CRADAs, will provide the framework for a set of data alliances led by each of the anchor companies. Data alliances, which consist of participating organizations across the private and public sectors, will work to research and test solutions for bringing the National Oceanic and Atmospheric Administration's (NOAA) vast information to the cloud, where both the public and industry can easily and equally access, explore, and create new products from it, fostering new ideas and spurring economic growth…"As America's Data Agency, we are excited about these collaborations and the opportunities they present to drive economic growth and business innovation," said Secretary Pritzker. "The Commerce Department's data collection literally reaches from the depths of the ocean to the surface of the sun and this announcement is another example of our ongoing commitment to providing a broad foundation for economic growth and opportunity to America's businesses by transforming the Department's data capabilities and supporting a data-enabled economy" (USDOC, 2015).

With this statement, the project was underway. The economic development aspects underpinning the announcement are striking. Phrases like the "goal of unleashing its vast resources of environmental data… supporting a data-driven economy" and "these collaborations and the opportunities they present to drive economic growth and business innovation," (USDOC, 2015) hearken directly back to the discussion earlier in this chapter of the assumptions that drove the creation of the project – promoting secondary use of NOAA data as an economic resource, increasing its use and – as a byproduct – dependence on it. As we have just seen, however, these dramatic aspirations are belied by the months of effort required to persuade the CRADA collaborators of the feasibility of the initiative and evidence of both their and NOAA's uncertainty about the commercial value of its data yet to be exploited.

And, as for industry, while there had been 70 responses, one might wonder what the RFI respondents thought about losing out without competitive bid, as an RFP was never issued. The project manager explained:

> [T]here was also a certain amount of blowback around "Why did you take this away from being an RFP? Why did you approach the infrastructure-as-a-service providers?" I would say none of it was particularly formal. We held various meetings and workshops and… It was more just sort of griping. (PRI-21)

It could be that through a combination of a) crafting an approach that made an agreement with potentially "agnostic" IAAS providers versus commercial users of the data – many of whom had included use of one of these providers in their proposals; b) requiring the IAAS providers to provide free access; and c) clear statements by NOAA that "nothing was going away" in terms of existing information supply chain channels, the respondents were willing to take a "wait and see" attitude. But, without a list of the respondents or more information on the business collaborations proposed, we are left with an open question of why industry's response was not more negative or aggressive. The next chapter sheds light on this issue and others encountered throughout this chapter as we see them echo through the implementation of the project up to the present day.

Conclusion

While this chapter tells only half the story (we shall see how the project progresses in the next one), this cradle-to-contract view through the eyes of the participants continues to build on the analysis of the impacts of dependence we've completed so far. In this case, much like the end of the previous chapter that examines the data promotion efforts of the NCEI, the tables were turned. Here, government decides to engage the market on their own terms – literally – with an idea for how they can benefit and industry can, too. The incentives are there: Industry demand for greater access to their data; an impending crisis in paying to hold, manage, and distribute it; an overarching policy directive from both the President and their parent agency to "liberate" it to jumpstart innovation and create jobs; and changes in the technology landscape for delivery that show things are headed toward a new paradigm for access to processing enormous datasets "in place." But, in some ways, the situation is like our very first case, Restaurant Inspections, but – again - in reverse. The government seems to have some ideas for how industry ought to behave (instead of industry developing an inspection standard for government to use) and why they should be incentivized to do so (they're apparently not, just as restaurant inspection departments had other barriers to participation). There is also some similarity in the misunderstanding of the role the organization and operational model of the businesses they approach (just like industry misunderstood government in the first case), as they describe going door-to-door, explaining the value over and over again. Who could imagine such a thing would be needed with the largest and most innovative technology companies in the world! There is also a similarity, even though this is open data, free for the taking, with the Driver History Records case, where we see government employees examining the market and attempting to calculate the impact of their decisions – in that case on pricing, in this one on industry adoption that would translate into significant cost-avoidance for data storage and delivery. We see again how complex those things can be to understand and predict, their volatility, and the difficulty of implementing a policy to address them, but this time constrained by the federal mandate for equal and open access.

Yet, there are new things here, too. Much like the last chapter with the NCEI – some of the government participants in this project, including its head, come from there – we see that all their actions are not necessarily reflecting the influence of industry. Some of the motivation may well have come from wanting to be responsive, but that doesn't explain, for example, why the distinct break between the EISWG/SAB recommendations and the efforts to start up a new and different project that didn't include those initiatives. In this case, it appears that government is attempting to solve its own problems with data delivery and be responsive to its own policies on opening up data, while ensuring they maintain control. But to do so, like with NCEI, they seem to take an entrepreneurial position, looking to "disrupt" (PRI-15)-or in the last case, create - the market, and are willing to change some practices, implementing a CRADA to have confidential discussions with industry about the value of their data, and proprietary ways they might design to access it. And, the recurring theme of market values occurs, with the implication that datasets that industry is interested in will serve as a proxy for the public, and democratic, expression of interest in what becomes public and what doesn't, again involving an archive. In this case, the dream is to expand the amount of data to more and more types to foster (and break down silos in) its use, but there is little sign in the provisions, here, for example, that NOAA will determine the priorities of what industry chooses to host – it is left to industry's entrepreneurial incentives.

On the other side, industry is dubious of the government's suggestions here, but we see a new value to them – the benefit of relationships, especially in light of other potential business opportunities associated with the multiple interests they have with government. The agreement is crafted so they really have little investment – there is no mandate to take any data, it is left to what they might discover in collaboration with this large government agency with a problem on their hands. At this point, the prospects might seem dim, but in the next chapter, the project and the envisioned dialog begins. We know from several references in this chapter that at least one dataset made it," NEXRAD Weather Radar. But was there a revolution, did NOAA get what they were looking for, and did industry finally see the light? Chapter 5 continues the story.

Chapter 5: The NOAA Big Data Project Part 2 - Making the Market Work

Introduction

As we left this NOAA Big Data Project in the last chapter, four large hosting and consulting vendors and one academic consortium had signed collaborative research and development agreement with NOAA to:

> …collaboratively develop a method for moving NOAA data from NOAA internal systems to a publicly available, cloud-based data repository and - provide equal access to all on equal terms and intelligently position the data near computing, analytic, and other value-added services, creating a new market space for economic growth and job creation. (NOAA, 2015)

However, it was clear that none of the parties had a good idea of which data they might choose to do this with – there was no obligation to choose any – and, to some degree, industry seemed to be going along for the ride, lobbied by a NOAA project team that played on their fears of missing out, and incentivized by, at a minimum, getting access to more people inside NOAA, something that might not hurt future business opportunities. In terms of our research design, this appears to be a case – so far – where NOAA needs to find a way to avoid current and looming costs of making larger and larger amounts of data available while demonstrating their commitment to openness and innovation. And, this is motivating them to lobby industry, rather than vice versa, and perhaps even do some things differently, to attract their interest. They've developed a model that they seem confident should make sense to industry, as it will provide them profit – a win all the way around.

In this chapter, we look at the fruit of that effort, primarily through a series of vignettes and interviews where we get to see two sides of the story: NOAA's hopes and concerns, and their conceptualization of industry's capabilities and motivation, and the perspective of private sector executives at very large, international technology companies – and their motivations – as the two struggle to make sense of what opportunities are here. At the chapter's end, the project is still in progress, but transitioning into a formal contract relationship. At its beginning, we wonder whether this will work at all, and see some obvious drawbacks to the model NOAA has designed, where hosting platforms, rather than weather or environmental companies who might use the data, are the ones charged with figuring out its value. We then find, along with NOAA and industry, that perhaps they don't understand each other's motivations – and capabilities – as well as they thought. The insights here may challenge our view of the private sector, but say more about the idea – as we've seen before – of promoting and grafting on open data efforts to organizations that were not built with that in mind. Both NOAA and the collaborators continue the struggle to find commercial value in new types of NOAA data, while traditional modes of making money, like selling the data, is prohibited by the terms of the agreement. But, even if mutual dependence does not develop, we identify yet more aspects of this phenomena along the way. Ultimately, it will be the customer's customers that use this data, and provide the funding to subsidize NOAA. And, NOAA finds that its expertise is part of the equation in making data valuable, something perhaps they have to provide, or trade, in order for the model to work. Overall, it seems – much like the efforts at the NCEI – to be a work in progress, and shows that, given the right incentives, government may be willing to make some changes itself to attract a suitor when a potential benefit comes into view.

From Idea to Operations: Initial Exploration of the Commercial Value of NOAA Data

The implementation phase of the NOAA Big Data Project presents a unique opportunity to learn about the evolution of mutual dependence on government data – it is, after all, the declared purpose of the project. We might start by looking at how CRADA collaborators work with NOAA to identify commercially valuable data: Did the companies arrive at the door with lists of data they sought to obtain from NOAA? Was there a joint exploration, and, if so, which party directed those efforts and how were decisions made? Or, did NOAA develop a clearer understanding of where value was latent in their data holdings and promote it? Did either party (or both) make concessions during this process and of what type? What kind of influence came into play – did either party attempt to build or use dependence – and, if so, how? Unfortunately, two aspects of the project make these questions difficult to answer. First, the CRADA requires confidentiality for some commercial matters, and actions taken by or conversations held with vendors are interpreted to fall under the category of business or competitive secrets that they may not, and NOAA cannot, share. Second, from the evidence available, the approaches used by NOAA and the collaborators change over time, something that does not seem unusual for a research project, but results in answers that are not uniform across vendors and remain in various stages of formation.

BDP Project Governance and "Data Prospecting": Ad Hoc vs. Institutionalized

The administration of the BDP is organized in a matrix fashion with most of those involved assigned to other full-time duties inside NOAA (PRI-18). The core team is composed of a project manager - at first the loaned executive from the private sector, later contract staff that have changed over time - NOAA's Chief Data Officer, and 1-3 employees from within NOAA that rotate in for short stints to gain professional experience (PRI-18). There has been no dedicated budget. There is also a small team of 4-5 stakeholders from across NOAA that share a history in the development of the project who meet on a regular basis to provide advice to the team in addition to their regular duties (PRI-24). Finally, there is a group made up of representatives from all of the NOAA line offices – business units like the National Weather Service, National Ocean Service, and the NOAA Marine Fisheries Service - that is considered the official internal advisory group to the project and convened on a bi-weekly phone call. Keeping these executive-level stakeholders informed on a regular basis allows them to serve as sponsors to aid interaction as needed with the sources of data within the offices they represent. As the head of the project explains:

> It's our opportunity to broadcast to everybody, everybody is invited - this is what we're doing, this is who we're talking to, these are our concerns. So, if we're looking for a certain kind of satellite data or something, well the NESDIS (National Environmental Satellite, Data, and Information Service – a division of NOAA) representative at the table can say, "Well, look, we hear that (the BDP team) are talking to so-and-so within our organization" and if they've got any concerns, they can bring them up there, they can bring them up later. Just making sure everything we're doing internally is all above board, letting everybody know what the communications are and what we're asking of whom. (PRI-18)

This description reinforces the team's work as a form of prospecting – there is no existing analytic or institutional framework in place for assessing the value (nor, in turn, relative priority) of NOAA datasets *as it relates to commercial use*. Contrast, for example, the process described in Chapter 3 (NOAA Observing System Integrated Analysis (NOSIA-II)) that NOAA uses for determining the value and investment priorities of NOAA's observing systems (and data), yet omits this explicit dimension – external parties use of their data - in the ranking process of determining which data gathering produces the most benefit to society. And, again, as noted in that chapter, there is a dedicated Chief Economist at NOAA, one of whose main roles is that of determining the value of NOAA programs and data. Despite emerging signs of change of incorporation commercial use into valuation work in those areas, these processes do not (yet) appear integrated into either planning or carrying out BDP-related activity. In essence then, there is not a solid, formalized, and shared understanding of the commercial value of specific NOAA datasets inside the organization, and no organizational processes in place to assess it. On its face, this seems likely to be the case with most government agencies who would dedicate resources toward optimizing internal, vs. external uses and value of data. The ability to leverage the commercial value of data to exercise influence in such an environment – either by NOAA, or on NOAA – is negatively affected by this lack of broad, common understanding of value by the parties.

"You tell <u>us</u>": Early Attempts to Identify Commercially Valuable Data

With this framework, the NOAA project team began to engage the collaborators to identify datasets to be moved to their hosting / computing platforms. While these discussions are interpreted by NOAA to be confidential under the terms of the CRADA, enough information has been shared by the parties involved to piece some of the story together.

The most significant finding from the early days of the project reveals a mismatch in expectations between NOAA and the collaborators that calls into question NOAA's assumptions about the strength of their entrepreneurial capabilities, or even interest, in identifying commercial opportunities for use of its environmental data. And, thus, for our purposes, the potential strength of mutual dependence that will be achieved here. We see this deficiency in the following description. While over time the ideas for potentially valuable data sets came from varying sources, the head of the project observes that in the beginning:

> [T[hey are coming from the collaborator's industrial partner and they approach us with a request. Most of our early discussions in this project came that way - *and most of the failures that we had were based upon that* **[**emphasis added]. And I'll tell you why I think that is, it's because of the perception that 70% of NOAA data aren't available. The folks that were coming to the table early on were looking for a unique opportunity: "Look, I'm a company, I've got this need and, uh, I'm going to partner with, say, Amazon and we're going to go get that NOAA data and we're going to put it out there." And they come in and talk to us about this stuff and they think they have this unique idea and, as we're talking it through, we're like "Oh, no, that data has been available for, like, you know, five years." And, in some cases, the partner didn't realize that they already had the data, there were a couple funny ones like that, right, like "Oh, yeah, we're delivering that data to you under some other agreement."
>
> So, that proved not to be maybe the best way forward and that kind of made us really

> stumble along for the first six months as we kept having failure after failure that way, *as we'd kind of made an assumption that industry was going to bring those good ideas to us and we'd be off and running* [emphasis added]. But, I think because industry was looking for - to be blunt - the not very creative things, that just the fact that you had THE only access to a particular product was their business model. And unique access in Open Data world is a sucky business model, right? And so they were all tripping over that same thing at first. (PRI-18)

Seeds of these challenges can already be seen earlier in the questions and answers in the second RFI (see Chapter 4), where companies repeatedly inquire about NOAA's data inventory with the expectation that NOAA had already identified (or would identify) commercially valuable datasets. When placed in light of the efforts by NOAA (recounted in Chapter 4 in the section "Competitive Incentives and the Known Unknowns") to sell the value of this collaboration, the reaction of the collaborators to the continuing difficulty in identifying commercial opportunities is understandable. One collaborator, for example, stated they were "flabbergasted" when they first approached NOAA after signing the CRADA, as they were "ready to go" only to find NOAA could not identify which of their datasets were most valuable, saying, essentially "You tell us…" (PRI-23) This anecdote again underscores a fundamental disconnect in NOAA's assumptions about the orientation and capabilities of the collaborators they had chosen as the foundation of their Concept of Operations and the workings of the market for secondary information. Their conceptual model saw the collaborators as incentivized to find customers that would use NOAA data in place and pay for the storage and computing cycles to do so, covering the cost of the free public access that NOAA required. But, as one of the collaborator executives stated:

> [W]e don't really know anything about your (NOAA's) data assets and who your community is and who your users are and things like that…*What we found is they didn't really either* [emphasis added]" (PRI-23).

In turn, one of the BDP principals at NOAA directly involved in the design and startup of the project reflected on the situation of the collaborators:

> [W]e've discovered that they didn't necessarily have knowledge of how to generate value-added services. So, they get stuck and they have a real hard time doing their piece of the R&D (research and development) because they don't actually have the domain knowledge. So, we find that instead of doing our own R&D…we're assisting our partners in their R&D to see if they can make money on it. (PRI-19)

Another member of the NOAA core team concurred, noting:

> (There's been) the recognition that there's this gap of people that need to evangelize the data and those processes….It's really turned into this structure of trying to educate these partners about what data we have out there, and explaining it, and trying to prompt ideas. (PRI-15)

In this case, of course, we can see that it is in NOAA's self-interest to help the collaborators find uses for the data to enable them to subsidize the cost of its distribution. But, as we've seen, it isn't clear that NOAA has a good understanding of potential commercial uses either.

"Pitching Data": Bi-weekly Collaborator calls

This move to direct involvement in helping the collaborators understand the potential commercial value of NOAA data is evidenced by the method used to engage them. Every two weeks, the BDP project manager and director invite a NOAA subject matter expert they have solicited to present on a conference call to representatives from the five collaborators (all are invited, not all always show up) (PRI-24). While the specific content of the discussions in these calls is interpreted to be confidential under the terms of the CRADA, some general information about how they usually proceed was shared by the director of the project and substantiated by the collaborators:

> The bi-weekly collaborator call, as we call it - that is our opportunity, again, to minimize the amount of NOAA time necessary to bring a new idea to the table…The main thing it has been good for has been to present new ideas. If there's some opportunity, some new data that's available, we'll have the expert come to the call, and they can spend just 15 minutes and reach all five collaborators and do this - we call it a "pitch," because it sounds like a sales pitch – "Here's this idea I have and here's this use case that I know about, and here's an opportunity for the data." So, they're pitching this data and they're

reaching all five collaborators.

> The calls are pretty funny in that the collaborators won't speak in front of one another - they're competitors and they don't want to tip their hand…So, it's deathly quiet. I usually have to warn the NOAA experts that are coming to give their little description that it's going to be just "crickets" after they make their presentation. But, typically what happens is that as they're making the pitch the collaborators that are interested in the product or the service that are being described by the NOAA expert will reach out to myself or the project manager, usually by text or email, and say "Yeah, we're interested in that, we want to do that." They won't say it on the phone, but they'll say it that way. Then, after that, we set up the one-on-ones. So, each one of the collaborators, if they're interested, will have a one-on-one with that NOAA expert so they can get into the details. So, that way, we're trying to be respectful of that NOAA expert's time and say "Okay, you're only going to have to sit and talk maybe a half hour or an hour with this person, only if they're really interested" - and very rarely are all five of the collaborators interested in a product, maybe two or three that he or she has to engage with, so we're trying to optimize our time that way. (PRI-18)

A sanitized overview of this approach is confirmed by the following slide (Figure 11) outlining the BDP "methodology" from a public presentation by NOAA on the project:

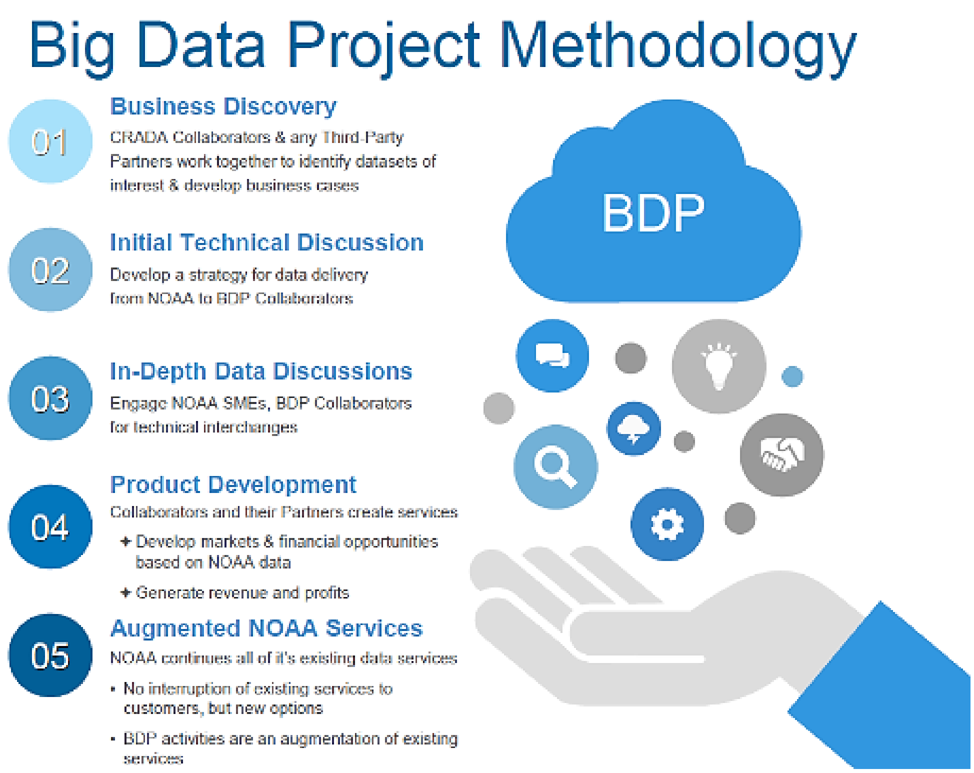


Figure 11 - Big Data Project Methodology (Casey, et. al, 2018)

Understanding and Promoting the Commercial Value of NOAA Data and Expertise

The interviews with NOAA management and the collaborators revealed little information about the datasets that have been promoted or rejected, or the mechanics of those decisions. One can see the end results only of those datasets that are accepted from slide decks about the project or from the collaborator websites. But, it apparently involves an ongoing process of soliciting and maintaining the collaborators' interest, as the project manager involved in the calls observes:

> We try not to bring things to the collaborators that are a waste of time. We're not going to bring them something that there isn't a public need or a commercial need for. So, we're typically bringing things forward that we suspect or have anecdotal evidence of at a minimum…These companies are making an infrastructure investment in the project, not to mention their time, so we don't want to waste that or be disrespectful of it. So, we kind of bring it to them and let them make the decision – if it aligns with their business case or not. Typically, what we're doing is we're saying here's the people we talked to and here's the business case and we'll bring it forward to the collaborators in a collective manner. We try to get a NOAA data expert on the line *to explain the data set as well as some of the business cases that they see for that, because we do have people that contact our archive and our dissemination folks to talk about, "Hey, my business is X. I need this data to do something." That is an experience or an insight we can bring to the table when we're talking with our collaborators about this (emphasis added). One of the limiting factors in that process is, with the exception of OCC (Open Cloud Consortium), the collaborators and the individuals in these companies that we work with are incredibly intelligent, but they're not scientists* [emphasis added]. They don't have the domain expertise and they're not meteorologists, they're not oceanographers, they're either computer scientists or data scientists. (PRI-24)

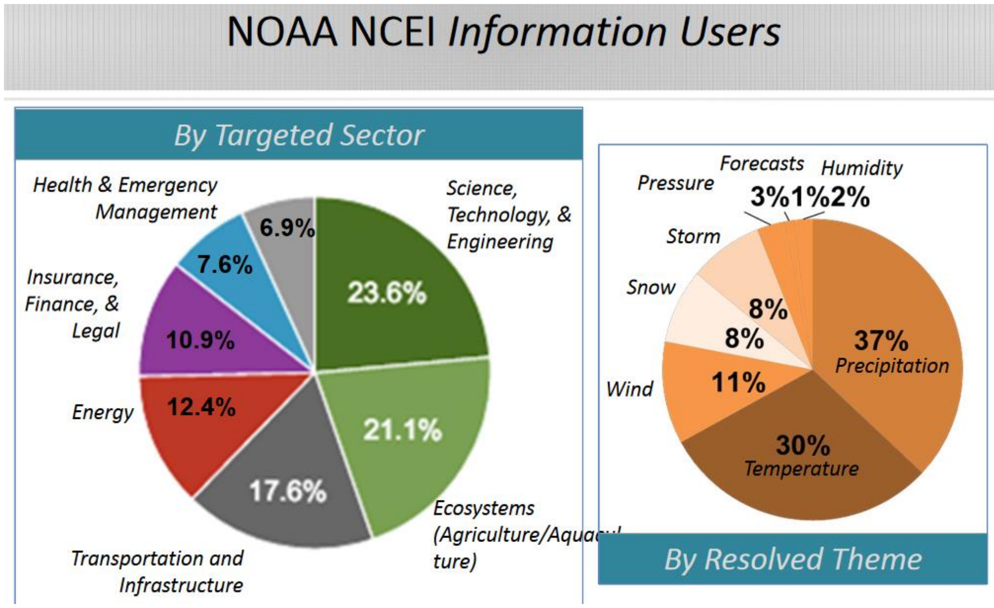


Figure 12 - NOAA NCEI Information Users (Kearns, 2017)

Two ideas are highlighted here. In reverse order, there is the concept – reinforced in previous quotes from vendors and principals in the project – that the collaborators lack insight into commercial uses of the data…and a conjecture as to why this might be: Lack of domain knowledge. In turn, another idea is presented, almost in passing, that NOAA experts are, at least in part, pitching the potential commercial value of the data <u>based on use cases (examples or models of uses) derived from questions or interaction NOAA has had with other commercial vendors</u> who have sought to inquire or use this data from the archive at NCEI (or via other means). See Figure 12 above, a slide from a NOAA presentation that shows two quantifications of data requests at NCEI. This latter finding is a different type of "value add" to the discussion than, for example, scientific expertise about the data like its frequency of creation / collection, quality, or uses internal to NOAA. While sharing scientific expertise in the traditional sense certainly occurs, knowledge about commercial uses of data gained by NOAA scientists and staff from observations or involvement with direct requests from industry is itself a type of data / knowledge. Of course, in this case it is knowledge only NOAA is in the position to gain and it is being shared confidentially with a small set of vendors. This situation could, in a sense, constitute a form of privileged access to business information to the degree that distribution about knowledge of uses was limited solely to them. One need only hearken back to the discussion in Chapter 3 about the government's concept of "Business Identifiable Information" (Privacy Threshold Analysis, 2018) to recall that one component was "[R]ecords that reveal commercial operations." While this might not be restricted when shared in aggregate or anecdotally, as we observe more about the conduct of the project and its increasing understanding of the value of NOAA expertise to commercial parties, the idea that NOAA acquires market knowledge simply through fielding requests for data is an interesting one that colors the discussion of value of the data itself. Note also, for example, such knowledge being used for programmatic purposes by the NCEI in Chapter 3 to "shepherd the use case" ("Making the Commercial Sector Legible") to foster the development of a business ecosystem around climate adaptation services. Ultimately, data has no economic value until one can figure who might use it and for what. While the rhetoric of innovation promotes this as a feat of imagination, learning what others have done seems like the most practical place for a business to start.

Pitching without Catching: Considering the BDP's Early Results

While lacking in specifics, these scenes either introduce or extend several concepts relevant to the exercise of influence by industry on the content, structure, and / or access related to commercially valuable data they obtain from government. First, previous chapters have established the context that – at least in the case of weather data – industry derives great benefit from NOAA, exploiting a market that they have negotiated through policy by fencing off NOAA's ability to compete. Yet, we see here that the commercial vendors that come to the table begin by looking for exclusive access that would create a niche necessary to profit, a business model even NOAA called "sucky." We are told that these efforts peter out without success. The occasional lack of awareness by companies of the source of data they use also presents a picture of the market as less sophisticated or effective than one might imagine. Second, NOAA's choice to create an experimental distribution model for data by placing at its center companies that – while some of the biggest names in business – apparently lack the organizational and strategic orientation, experience, and ecosystem ("domain") knowledge needed to profit from the arrangement appears to bear some responsibility for this situation. Of course, this design was a conscious choice we see being made in Chapter 4 as part of a strategy to "disrupt" the information market (PRI-15), and so it has. The signs of this problem are clear in the questions and difficulty in persuasion encountered by NOAA during the selling of the model to the collaborators seen in the last chapter. And, the examples above show these same issues continued unabated once the CRADA had been signed. As a result, the hypothesis that industry would solicit, then begin to depend on, new types of NOAA data commercially and, in turn, exercise influence that would affect the content of the data, its structure, and / or access and distribution policies is not yet proven from what we are seeing. Third, the value of government's understanding of the secondary uses of its data emerges as a factor in mutual dependence and the exercise of influence in government data sharing. In this case, at least, only NOAA has that knowledge and they appear to be trying to use it to create / increase this dependence, to the point of offering "suggestions" (see next section) to "help(ing) our partners in their R&D to see if they can make some money" (above). There is some similarity here to efforts in Chapter 3, where we see a program associated with the NCEI capitalizing on the same knowledge to further explore markets for its data in furtherance of its mission to encourage private sector investments in climate adaptation. In turn, it should again be noted that – other than this knowledge – NOAA appears to lack an institutional mechanism for assessing the commercial value of its data, which impedes their effectiveness in providing this help. Finally, a complementary and quite plausible explanation fueling these challenges also involves a choice made about the model: that the project is labeled and seen as experimental. The market for NOAA weather data is already well-established and there is little incentive to switch to a new model until it becomes financially attractive (which requires some level of reliability), especially given that NOAA has committed to continue to make it available via existing mechanisms. As we see the project finally discover a dataset, however, 1) the model has not changed, 2) the pitch did not originate with NOAA scientists, and, 3) once again, the collaborators had to be sold on its value. It is one of the Presidential Innovation Fellows, a former weather executive on loan to government, who brings forward an idea long known to him from his previous role.

Something Old, Something I Knew: Marrying Past and Present in Establishing the Commercial Value of the NOAA Next Generation Radar (NEXRAD) Dataset

Industry Knowledge and Identification of Data (Part 1): The new idea…

In the last chapter, we see the BDP team using a government program to hire two individuals from the private sector for a short-term assignment on this project. In this case, the hiring coincided with the timing of the RFI / Industry Day for the purpose of enlisting more help with the BDP. However, one of these individuals had some previous experience in the weather industry. As we have seen, once NOAA had made the decision to focus on the IAAS providers - now “Collaborators” - as the foundation for their concept of operations, there was continual back-and-forth, both parties struggling to identify the first set of data that could be commercialized as the clock ticked. The director of the BDP picks up the story from here:

> And then the NEXRAD* thing got moving. The NEXRAD idea came about - it was very fortuitous - it was sort of an ideal dataset for this because it was a really hard dataset to use from the NOAA system because it's like 200 million files, 300 TB in size, on a tape system. *And, there had been in the past a number of very smart, very good researchers and companies that had tried to extract all of that NEXRAD data from the archives at NCEI - and they had failed, they had all given up* [emphasis added].
>
> One of those guys who had failed, OK, was on our <u>team!</u> (The loaned executive). So, he was able to <u>personally vouch</u> for two things. One, the value that this dataset (NEXRAD) had for the meteorological services industry, and two, that under its current status in the NCEI archive, it was not useful. And, so he was able - and when (one of the collaborators) asked, and this is probably almost a direct quote from the woman that was working there at the time, she was like, “You know, but is this really worth it?,” - he was able to personally vouch and say "Oh yes, it is" and explain why. So that was really important to get that ball rolling.
>
> And, I think, for the early days, all the datasets that moved, that got over the hump, had at least one identified commercial user that enabled the collaborator to say "Okay, look, I think at least I can cover my costs." I'm making an assumption there, but they had enough motivation to say “Look, I know the use case, I can justify this to my bosses or whatever, okay, let's go forward.” And in that case, Climate Corporation was that commercial case that was willing to step forward and help with that and the partner was Amazon. But, of course, as we learned, all these other users came in and users followed Climate Corporation in and continued to follow them in to the Amazon environment to work with the NEXRAD data. But those other users really weren't identified, only that one big user was, but that was enough to get the motivation going. (PRI-18).
>
> * NEXRAD, or Next Generation Radar is the data behind the familiar weather radar images, widely available to the public that are often used to identify arriving storm fronts.

From this account, the selection of NEXRAD data did not originate with the collaborators, nor NOAA, but its value was identified by an executive on loan to the NOAA BDP team who "sold" the idea to the collaborators based on their confidence in the legitimacy of his experience and knowledge of the commercial possibilities. The influence here, again, then, does not seem to come from industry in a direct way, but instead, via use of industry experience by government to further its purposes.

Industry Knowledge and Identification of Data (Part 2):…was an old idea

Or perhaps it is a little more complicated. The choice of NEXRAD is “fortuitous” (PRI-

18) from more than just the Big Data Project's perspective. For it turns out that the BDP is but the most recent step in a long history of interplay between content, structure, technology, access policy, distribution methods and commercial influence surrounding NEXRAD data that first began with the advent of this technology in the 1990's. The next chapter explores this story in more depth and - taken together with the initial decisions and actions around the selection of this data recounted below – uses that history as a lens that helps place the choices and exercise of influence in the BDP related to this data in a larger context, revealing elements of path dependence and co-evolution of changes in technology and policy that will help us better understand the drivers and constraints that shape dependence and influence in government data sharing with industry. For now, however, we move to the perspective of that loaned executive on how things played out in the BDP, an approach clearly informed by the past:

> So, (when he worked at a weather company) a bunch of us had this great idea that we'd like to have historical radar on our map product and be able to go back in time and see things like Hurricane Katrina….So, we started downloading the old radar data, and we started saving the new radar data. And the nice part about saving the data was every day we'd have another day of archives (and) as I'd like to backfill it with data from NOAA, I spend about three years writing a script, buying hard drives, popping 'em in to a random computer I had, and downloading terabytes of radar data from NOAA…And it took three years. It shouldn't have taken three years. My day job was running a company. Maybe it would've taken about a year, probably not much faster than half a year. But the scripts would break, something would change, and we were just totally doing it without asking, just going through public interfaces**.** *And so that data's always been available, but it was on a tape archive, and it was really slow and hard to get to and hard to access. And, so, I think that's one of the interesting things about BDP was that a lot of its first data sets were not controversial…They were out there, just the access wasn't great* [emphasis added]. (PRI-22)

> So fast forward, after we sell our company to another weather company and I was working there I met the BDP team without realizing it when they came out to do a tour and promote and try to get the company to sign on to this project and be excited about it…And then six months later, I see a job posting on TechCrunch. I said "Well that's interesting – and, wow, that must be that project I heard about, and that seems great. I'd love to help them do that." (PRI-22)

In this case, then, we have a private sector weather industry executive and entrepreneur who joined the project with the knowledge of the commercial value of at least one specific dataset, along with deep experience with that data. After the executive signed on to the project, he became distanced from the contracting part (PRI-22) which was led by the other loaned executive. Instead, he became:

> ...[T]he weather person, saying "Hey, this data set is useful or that data set is useful. Then, having known that I really, really wanted the radar data (NEXRAD) for myself, I figured a lot of other people did, too. Then at the time, (the individual who was later project director) was at NCEI in North Carolina where they had all the data. And when I chatted with them, they knew what I knew. We were talking the same language, which is "Oh, everyone uses, loves, wants the radar data," right? And they had stats, which was the great part. They had data (on usage). They had numbers of how many people downloaded it, and we all felt like the numbers were already huge. Our intuition was given how hard and slow it was to get the data, if we made it easily accessible, it should just be even bigger. And we wanted to do satellite and then a few other things ... And then the problem became that a lot of industry wanted to do new datasets that didn't exist. And I would love to do, or we all would've and hopefully they will, and they did do after I left, new data sets that didn't exist. Like they did the GOES-16 data, GOES-R, whatever you want to call it (Note: Geostationary Operational Environmental Satellite). (PRI-22)
>
> But, at the time it was a little frustrating because there's two issues, right. One is it's very easy from the NOAA standpoint to get buy-in of "Oh look, we already have this data public. Let's just put it out public a different way. In the scheme of the world, *there's pushback (by NOAA) in "Is it going to cost more money, are we going to have to support it, can you ever turn it off?* [emphasis added]." But other than that, it's not that big a deal. But if you pick a new dataset, then you end up operational, potentially, because people are going to start using the data operationally" (PRI-22). [*I*]*n case the CRADA didn't work, and things needed to be rolled back, it is hard to take back services and turn them off once you start them* [emphasis added]. (PRI-22a)

Conflicting Commercial Interests and the Value of Data (Part 1)

From this description, it is notable that while Open Data furthers NOAA's mission by extending its accomplishments and complementary activities outside its organizational borders, it is not without cost to create and distribute. So, there may be an organizational disincentive to devote resources to this creation and distribution, i.e. a new "service" that is not a specific scientific mission, as opposed to some other mission-related purpose. In turn, the motives of the commercial users of the data may also conflict. In reflecting on what happened, he continued:

> And on top of that, maybe they didn't want the data to get out at all. So that became a big issue where a lot of the private companies kept saying "No, that's not useful." *And it was wild having been at (one of these companies) and using the data, and then coming out and them telling me it wasn't useful. I was just like, "I don't know what you're talking about. It was useful when I was there* [emphasis added]." (PRI-22)

He explained his thoughts on the rationale during this and a subsequent exchange:

> I think partially they didn't "get it," and I think partially they felt like they already had it or they had some way to do it. They didn't want competitors to get it easily, too…As soon as they thought about old stuff they'd already done, I think they didn't want to go back and redo it, and they thought "Well, that's just going to help other people, not us." (PRI-22)

> [I]f you were AccuWeather and (The) Weather Company (etc.) you had already spent a bunch of money on special connections to NOAA, satellite dishes, etc. They didn't need to redo any old stuff, and in fact they didn't want to spend money doing it. *Plus, it created a "moat" for their business*" [emphasis added]. (PRI-22a)

Here, of course, we see competitive aspects and path dependence that weakens the ability for NOAA's program /distribution policy implemented through the BDP to reach its goals. The biggest companies had already made a good enough investment for their needs. And, in the case of the weather industry, the size of secondary players drops off significantly from the top companies. Of the approximately 199 companies involved in this work in the United States in 2016, IBM (having purchased The Weather Company) and AccuWeather account for roughly 67% of the revenue share, with none of the remaining 197 companies accounting for more than 5% (2014 *IBIS*World Industry Report, as cited in Longden, 2017).

Exploiting the Incentives of Industry

Just like there had been an attempt to play off the vendors against each other to get them to join the CRADA, they used this approach to sell collaborators on interest in hosting the NEXRAD data. He explained:

> Basically what we were trying to do with this whole partnership was get a "pull." (So) we were trying to get this pull from companies - from The Weather Channel, from AccuWeather, from everyone, for them to tell their cloud provider ("collaborator" in this case) "Hey, I really want this dataset." And then the cloud provider would say "Oh, okay, two or three companies are saying this is important," and then they would sign on, saying "Oh, we don't want to get into a situation where only Microsoft is in and we're not." We had this huge game of chicken where we wanted all of them to sign up. (PRI-22)

While NOAA played the collaborators against each other to get agreement on adoption, there was another factor shaping the approach. In the loaned executive's words:

> So, me coming from industry, one of the things I didn't want… the FAA used to do a lot of projects where they would give or sell the data to one company that would then make it available to everyone else at a fee. And the original radar data back in the early 90's or whenever, there were two or three companies that had to pay to get the data and were selling it to everyone. And (our company) could not really afford radar data for the first five or so years of our company's existence. (PR-22)

This portion of the history of NEXRAD distribution is addressed in the next chapter, yet it is obvious that this influenced the approach this executive advocated for and that he saw advantages to finding a distribution model that would keep the data free. To him, this meant hosting by multiple providers (Collaborators).

Industry Choices about Technology Shape Equity

But, there was yet another reason that could be used to leverage this competition that was related more to the mutual exclusivity of collaborator platform (IAAS) <u>technology</u> versus <u>exclusivity of access to the data</u> – although, in fact, they turn out to be related. He continued:

> So, I kept looking at it like okay, if I'm back at putting my startup or big company hat on, either one, what am I going to need from NOAA? So, I kept thinking well first of all, it would be really horrible if NOAA picked one cloud provider. Because then if you're with another cloud provider ... My example is at the time, The Weather Channel was in Amazon and AccuWeather was in Microsoft. And so every time I would be in a meeting with AccuWeather, I'd say "Well you don't want to be in Amazon." And every time I was in a meeting with The Weather Channel, I'd say "You don't want to be in Microsoft, right?" 'Cause everyone talks about the cloud as this one thing, *but if you're in the wrong cloud, that's not really helpful* [emphasis added]. (PRI-22)

While this technological dilemma may seem complex, the premise is straightforward. NEXRAD, or other, data may be formatted using independent standards, but the processes and tools used by the different collaborator hosting (IAAS) environments are different, each requiring specialized training and skill sets. If a company (in this case, a potential member of a data alliance) has heavily invested in storage and processing of data on Amazon, for example, it adds another significant cost – and risk – to the equation to either switch to another platform like Microsoft (perhaps for much more computing work than just this) or support two different platforms. In turn, having everyone agree to participate at the same time wouldn't disadvantage any party and, from a practical perspective "[W]e didn't want to transfer the data to one provider, and then go back through and do all the work again to transfer it to the second provider" (PRI-22a).

Where We are So Far: Industry Matters

With the initiation of the NEXRAD datasets as the first major BDP initiative, we start to gain insight into a microcosm of the influences that can come into play. So far, we have seen a potential mismatch in understanding between the organization creating the data and the market focus of the company(ies) that are used to host it: marketing NOAA data and identifying potential customers are two things they are not set up institutionally to do, nor are they familiar with the data or customers. In cases where their customers may be familiar with the data, to the degree that they have already made investments in getting good enough access to it, they may see further investments that improve access for themselves, but also benefit other companies, as against their own interest. Equitable access is a tool to be used to aid their competitive position, but there is no incentive to support it if it degrades that position. Then, the fact that the commercial data users had already picked different collaborator platforms (say, Amazon vs.

Google) prior to the BDP shows how path dependence in the external information market can limit choices and forces government to consider multiple platforms for distribution in the name of technological equity. Finally, an executive from industry – loaned to government - who had experience with a NOAA dataset and saw a specific business case with a potential return on investment that would have benefited him when he was with one of the weather companies was seen as coming up with a new idea for targeting NEXRAD data. In actuality, then, there was commercial influence on the decisions in the form of a) the experience and knowledge of an executive loaned to NOAA, b) his recognition and anticipation of pre-existing aspects of the commercial market, including the lock-in and path dependence that allowed NOAA to play a "game of chicken" (PRI-22) to incentivize collaborators to host the data, and c) the resulting shape of how and where they delivered data to conform to those commercial investments. This knowledge comes, in turn, from his experience with the history of NOAA data distribution choices with this specific data.

The Collaborator and the Customer: Industry Perspectives on NEXRAD and the BDP

Influence and the Customer's Customer: Climate Corporation's Interest in NEXRAD Data

Conversations with the community of commercial users of NEXRAD data support and expand this picture. Climate Corporation ("Climate"), a unit of Monsanto, focuses on precision agriculture, delivering customized data to farmers to help them better understand – and anticipate – the impact of weather on their crops (Climate Corporation, n.d.). They were the principal "customer's customer" that created the "pull" referred to above for the NEXRAD dataset to be hosted by Amazon. A closer look at this helps frame the chain of influence operating outside NOAA that eventually touches them.

As an employee of "Climate" directly involved with the deployment of NEXRAD in the BDP noted - along with weather forecasts and observations from past weather noted "(One) of the critical things for agriculture is knowing how much rain fell and where" (PRI-25). But, unlike the public forecast, Climate provides this at the agricultural field-level. Their interest in NEXRAD data came from feedback from their customers about the quality of the data in an app they had deployed that included precipitation data. Up to that point,

> We were purchasing products from a third party and basically just extracting the location of the (agricultural) field and providing that data. There were issues around how we were using the data and the quality of those datasets. We decided to take a look and see, "Could we do it better?" (PRI-25)

While various options were considered, including other external sources of the data, as well as working to add value on top of purchased data before it was delivered as a product, one option was obtaining data directly from NOAA. There was an added incentive for this approach in that Climate now had an expert on staff that had worked directly for the government for many years with the NEXRAD data on hail and precipitation analysis. Climate was already accessing the data through the existing methods NOAA had put in place. However, echoing the earlier observation about difficulty of access by the NOAA PIF who had worked with the data in the private sector, the Climate manager stated:

> It's freely available, NOAA provides it – however, it is extremely difficult to get access to. You would have to put in a request, fill out a form, a machine goes and pulls the tape, uploads the tape, then they send you an email. Then you have to download the data. You can imagine that process taking months to do an evaluation of the entire country. The challenge was that we had tons and tons of radars and data and tapes that we were trying to pull. You end up that one tape is harder for the machine to find or whatever it's in the queue for reading process, but then you don't download it fast enough, then you lose it, you have to go back and then make sure that same data was retained in full. (PRI-25)

Here we see the impact of government decisions about technology and policy – and the consequent path dependence of these choices - on how (and what) data is distributed to commercial users and its intersection with a commercial need - that is, the information market. This manifestation is directly linked to the challenges identified in Chapter 3 related to the inadequacy of NOAA technical infrastructure in supporting a private sector supply chain. Most importantly, this effort reveals how the channel was opened for industry to influence government to make data available in a form that was more useful – and more profitable – for them:

> It's something that people want. But, people don't really want to download all of it, they only want to have access to certain amounts. The people who really want to use it don't have access to enough of it to make it useful. So, we made a case with NOAA as well as with Amazon. With Amazon, it was they were more than happy to do it because storage costs for them are fairly cheap, it's the compute cost. For us, we've stored everything in Amazon. (PRI-25)

Sources of Subsidy: Funding the Commercial Information Supply Chain. This statement also reinforces the propositions just identified with regard to sunk cost (path dependence) and technological equity in distribution. Climate is already using Amazon for everything else, so having the "free" data there is much different than if it were only available through Google or another platform that would come with additional costs. In turn, we see that in NOAA's attempt to structure its data distribution through intermediaries, the cost and provision of what I will call the "distribution subsidy" to NOAA is borne by the "customer's customer" – not the collaborator (IAAS provider). The manager at Climate continues:

> Having access to the data means that we can do more larger-scale evaluations. So then we're doing (incurring) more compute costs. *And so, we're essentially helping them pay for the data* [emphasis added]. As far as I'm aware, as of when I was involved...Amazon was the first and the only one that really had the data available for the NOAA Big Data Project *because they had a customer, that customer being us* [emphasis added]. I think for us it was a combination of things that allowed us to really take advantage of it. One, it was the infrastructure we were already heavily using and that was really a benefit. I think that's why if the others don't have the data up and available right now that's why they don't *because they don't have a partner who is willing to be the one to help, if you will, subsidize the data* [emphasis added]. Amazon is making money off of our compute costs. Not only on research side, but on a daily basis in real time. That's helping to subsidize the cost to store that data. If they can get other people to also use that data now they're starting to generate a profit. (PRI-25)

Conflicting Commercial Interests and the Value of Data (Part 2). In an earlier section, we saw that some larger companies find their competitive position served by the status quo. Having made investments to compensate for NOAA's shortcomings, they were less interested in a level playing field. However, in the following story from Climate Corporation, we see another dimension of their calculations – and how it may differ depending on where a company sits in the information supply chain. As part of moving NEXRAD data to the Amazon platform, Climate Corporation discovered some errors in the data and helped in their correction (Ansari et. al, 2018). The correction of errors is obviously different than, say, materially either shaping the content of the data or its structure, but looking behind such a straightforward term ("correction") reveals tensions and decisions that could play out differently in other data sharing scenarios. The manager from Climate explained:

> I developed code to check Amazon's archive compared to what NOAA said and we were able to find a long-standing bug in their data where there was some corruption of files that they were able to go back and regenerate and have a much more complete archive. We never would have found it without this project because that's not how their people are typically downloading data. Like I said they're downloading a radar at a time and it wasn't just a single radar problem. We provided them with the guidance that "Hey this is an issue." They dug into it on the NCEI side and said, "Oh yeah, thanks, you found something" and then fixed it. [I]ndustry and private entities have a different data use but there is a potential that we could just hold on to it and just fix it on our end or we could let them know for the benefit of everyone. Our point is if it's not large intellectual property (IP) value for us then we're going to help improve the data. We're still scientists by nature and we reached out to them and they were able to fix it. Depending on the dataset that may or may not be something that a company is willing to do, but for this what we're doing ultimately with the data from that raw data to get to what we're actually using it for there's a lot of secret sauce that's IP - the raw data is not IP. (PRI-25)

It is important to note that – even between collaborator and their customer (the end user) – the motives and attitude toward roles and responsibilities may differ. Another take on the story above, this time from a representative from one of the Collaborators:

> In the case of NEXRAD, for example, there were gaps in their archives that they didn't even know about, but their initial response when we found out that there were gaps in the

> data, they're like, "Hey, so there are gaps in the data. You've got to fix those." We're like, "We don't care about your data." That blew their minds. They're like, "What do you mean you don't care?" Because from their perspective, and again, I think this is fair for them to have thought this way, but they were like ...They had just given us this amazing asset that we were going to productize. That we should be really excited to have it and it was going to make us tons of money, so we should want to clean up the data. We're like, "No." I characterize how resources are used in terms of risk, right? For us to allocate resources to cleaning up NOAA's data is a risk, right? Because those resources could be used for anything else. For us to say, "Okay, we need to spend X amount of hours cleaning up this data because we might make money off of it," we just didn't see any evidence that that was really true. (PRI-23)

Even if there is an element of hyperbole here, the account rings true. In the case of the customer who had the use for the data, they stood to benefit from the corrections. An intermediary, while seemingly interested in making available an attractive product, was really subsidizing the availability of NOAA data from Day 1. In business, with many alternative options for investment, the return on every project is likely under constant review. Both these anecdotes speak to the complexity of motives and interests in an information market and the challenge of crafting government policy that would shape, much less counterbalance, incentives for influence.

Government Resources are required to satisfy the Demands of the Market

This observation is important to the study of influence. The demand for data in an ecosystem where dependencies among external parties already exist means that when government seeks to cultivate a dependence, the visible face of the dependence – in this case, Amazon – may sit in front of many layers of other users…such as Climate Corporation and, in turn, the customers who depend on the reuse of this data downstream through Climate's (or others') services. This effect is acknowledged by NOAA as a goal from the perspective of economic development / job creation and something they sought to leverage in realizing the value of the data – the untapped use - they refer to as "latent" (PRI-18). And, they are aware of it from their history with distribution of weather data discussed in Chapter 3, where they incorporated policy provisions at industry's urging to reassure them that all of NOAA's decisions will carefully consider potential disruptions to the existing information supply chain (NOAA, 2006). However, the breadth of the professed scope of the BDP and its focus on responding to commercial demands for data anticipates significant growth in this dependence and, in turn, an expansion of this commitment to continuous availability.

At present, this is not the case. More than one collaborator discussed the dilemma introduced by NOAA's "best effort" (no guarantees) approach to providing data in the CRADA. As stated by one of the executives from a collaborator with regard to expanding their participation:

> And then, the other thing that came in that was a very complicating factor is, when you're looking at doing something with the data, your end customers - and I could extrapolate this to pretty much anything then to service-related business: *Your end customer wants performance levels* [emphasis added]…So, we can say, "Yeah, we can do this for these performance levels," but if we're dependent upon NOAA to source the data, now we have to have those performance levels rolled out contractually.
>
> So, if we had a supplier in a supply chain, and we said we were going to give you 10 bottles of water every two days, we would go to our supplier and say, "You have to agree to 10 bottles of water every two days, or I'm not going to commit to my client." And if they say, "Well, no, I might be able to, but there's days where I might not," then we kind of go, "Okay, well, wait a minute. I can't make that commitment anymore..."...... All of a sudden, if I'm needing that data, if I'm doing something with the data that's going into an app, and it's something that an organization or a community is depending upon, which is a life or death decision, and I can't meet my commitment - on the corporate side, now you're starting to talk about indemnification and limitation of liability. (PRI-20)

These statements further reinforce the tension between the design of a government agency's programs and process / technical infrastructure and the demands for external use of their data that they can reliably support – along with the implied tradeoff in internal resources (budget, manpower) that would be required to do so. A tradeoff between limiting technical and human investments to focus on primary internal use of data vs. larger investments required to support reuse affects the level of reliance and dependence that can be achieved by industry and, consequently, establish limits to use, dependence, and the incentives for influence.

Path Dependence vs. the Experimental Nature of the BDP Revisited

There is also an element of path dependence working against this experimental aspect of the project for longstanding users of real-time NEXRAD data. As noted earlier, these users can continue to obtain the data through previously established means, although these, too, may not offer guaranteed performance. A representative from a company that provides both commercial and consumer weather services sheds light on this:

> [A]s a private sector company, we get data from many, many different places and part of our goal is to have very high uptime, of course, and so even though our data source might be NOAA, we typically might have a satellite feed from them, an Internet feed from them, and might even have another satellite feed or Internet feed from a different place so that we can get triply redundant access to the data. Especially the important data like warnings and those kinds of things. Even radar data, we have a Level 2 feed from one place along with internet feeds from three other places, so that if the data's not coming from the Weather Service, we have some way we get to it…But I think in general, what we're finding is that the Weather Service, even though they don't guarantee it in an actual way, their stuff is up a pretty high percentage of the time. (PRI-26)

In contrast, with regard to NEXRAD via the BDP, he continued;

> We don't have the SLAs (service level agreements) from AWS (Amazon Web Services) or anyplace else saying it'll be up 99.9% of the time or anything. And some of the ways they do things aren't stable enough, or at least guaranteed stable enough, that today we aren't using the data from those places, honestly. (PRI-26)

Here, besides sharing a similar concern on the need for reliable access expressed previously by other vendors, we also see that this issue is handled in the existing "pre-BDP" world through an investment in redundancy versus service guarantees from NOAA. It also seemed to this vendor that the storage and delivery of NEXRAD data on the platform had been architected with a different use in mind, say, historical analysis, vs. real-time use. When the company approached Amazon to structure the NOAA data feed to be used real-time, they had reliability problems, which they attributed as follows:

> [T]hey (Amazon) don't look at it as an operational kind of ... They don't have an SLA on it or anything like that. So, they look at it as it's sort of a pet project that there's a lot of interest in, but they're kind of doing it out of the goodness of their heart at this point. There's no real funding for it. So, in talking to them, we made the choice that it wasn't probably something we were going to rely on. (PRI-27)

While this may change as the project progresses, it re-emphasizes the layers of dependence and priorities of commercial users that sit downstream from the original data and, in this case, the path dependence that would cause the still existing, well-known distribution methods to be good enough to reduce the incentive to press for a new alternative. This idea of path dependence based on previous investments is a common constraint that can limit the incentives for influence produced by new policy (Pierson, 1993).

Questioning the Assumptions underlying the Data Alliance Concept

Interviews with industry also added perspective to some of NOAA's other assumptions about the operation of the information industry ecosystem. The company interviewed above, in addition to providing both commercial and consumer weather services, operates their own information marketplace," brokering weather data to third parties for commercial uses. From that experience, they have reason to question NOAA's assumption about the likelihood of the formation of Data Alliances of users of raw data tied to each collaborator:

> They (NOAA) have this vision that there are data scientists out there who are going to say, "I need weather data," and they're just going to go say, "Hey, I know that Amazon or IBM has all this weather data online, so I'm going to go grab that data and I'm going to work with it." Well, that's conceptually not a bad idea, but I'll tell you what we have found just because (our marketplace) is marketed to data scientists, right? Data scientists say, "I don't want to spend the time to write a bunch of code to learn how to process that data. I don't want to learn, you know, that radar data is in radial data and it's in, you know, multiple elevations and it's radar-by-radar." They don't want to learn the details of the data because they have their own data that they have to learn the details for. What they want is they want weather data. They don't want technical data that they have to then turn into information, right? (PRI-27)

However, for the collaborators, especially given their position in the supply chain as primarily hosting the data, not packaging it for end users, this lesson may be new. It is notable that - without recognizing it as such in the conversation - the principal manager from one of the collaborators in the in the BDP began to describe something like a data alliance they are developing around GOES-16 satellite data (discussed later in this chapter):

> One thing that we've found in parallel to the CRADA has been ... It's actually come out of our work with the satellite data is that *we've created our own internal groups basically* [emphasis added] ... We've identified our own customers that are doing cool stuff with earth observation data. It's basically a mailing list - *What we found, actually, is that the best way to identify these valuable data sets is to create a subset of our customers who actually have opinions about the data. And that's where we start getting some traction. We as a cloud provider, we don't have any opinions – we're kind of agnostic. But we can convene a number of our customers who do have opinions, and so we've been doing that community organizing on our own, and I do think that the future of the CRADA is going to involve more work like that, where it's us basically kind of refining our audience* [emphasis added]... It's hard work, but doing the work to find who the people are who actually know how to use the cloud and see the benefits of it…It's not that many people. (PRI-23)

Magical Thinking: What can Intermediaries do, what does Industry Want?

One executive at a large commercial weather firm pointed out that a key to using the cloud approach was not just in the advantages of cloud technology, which removed problems of access and security from NOAA's sphere and provided relatively unlimited computing power for analysis, but in the amount and types of data that were placed there. Was it really opening up access?:

> [I]t's only making good on the idea if by that process, we're making materially different and new data available to the end-user that they wouldn't have had access to before. I'm not sure that that's true. And I don't see the CRADA on a path, where there's going to be 10- or 100-times greater access, data access via the NOAA Big Data Project than what we have today. (It's) not trying to get 50% more data access. We're trying to get thousands of percent greater access. So, "Yeah, yeah we made twice as much data available"? Don't even waste my time saying that. You know, it's got to be a completely different paradigm...By getting the real-time data into the cloud, they didn't make any more data available. It was already available to me and I was already using the data. I can now just use it in the cloud without just making a copy of it myself. But *it didn't solve anything. It didn't move the rock forward in terms of real-time weather data activity* [emphasis added]. Where it did, however, move the rock forward, is there are archives of

> these data that had been stored on, for the most part, magnetic tape, locked away in vaults in the National Climate Center (Note: Now called NCEI). (PRI-11)

From this one may take that, because it is an experimental project, there are not enough new types of data yet being hosted to generate the sort of dependence / use that might begin to produce influence. However, it also suggests questions about the feasibility and effectiveness of the current model being used in the CRADA and whether it can reach the magnitude or “critical mass” needed to satisfy industry. He continued, setting the problem in historical context:

> (With) traditional data, they haven't worried about this problem. They just said here's the data that we're publishing. Come and get it. They don't really care. One person's getting that piece of data and 10 people are getting that piece of data. There's no need to do that kind of accounting. But now if you asked the private sector to build those infrastructures for you, you're darn well right, those accounting are going to get done. That's where the CRADA bottleneck is. (PRI-11)

Similar observations are made about the bottom line and the feasibility of the model were made in an earlier section that noted the debate about whose responsibility it was to address errors found in the NEXRAD data. Of course, there were questions about the financial model – proposed by NOAA, not industry – from the very start. This uncertainty seems to be an obstacle to creating the dependence necessary for the collaborators (and their customers) to really get “skin in the game.” But, it also calls into question what “skin” the collaborators actually possess. Witness comments by one of the key collaborator executives reflecting on the situation:

> I think there's a bit of “magical thinking” that was sort of going around DC at the time (the CRADA was developed), and this persists, this magical thinking. It's not unique to NOAA or to the people involved in the Big Data Project. To simplify this equation, it was very much like “NOAA's data plus commercial cloud providers equals money for the commercial cloud providers, so that NOAA doesn't have to pay to distribute this data” - It was wildly optimistic, right? I think they came to us with very optimistic expectations that we somehow were a conduit to commercialization of their data. They have this huge set of datasets and they're like, "*Okay. The people who are going to tell us how to get value out of our data sets are these innovative cloud companies because they do stuff with data”* [emphasis added]. That was just from the category they used to define "These are the organizations that are going know what's up." Our response to that is, "No. We provide infrastructure. We don't have opinions about people's data.” We just want to

> provide services for people to work with it. The more the better for us. I guess it's like, if we have an opinion, it's just that we want more. (PRI-23)

Built for Something Else: Structural Barriers to Commercial Influence in Implementation

In the last chapter, we heard first hand of NOAA's difficulty in selling the BDP concept of operations to the Infrastructure-as-a-Service (IAAS) providers (now "collaborators.") And, that a significant barrier to their success in getting vendors on board was that none of the companies were structured appropriately (no customer-facing lead "intake" for such a proposition), nor were their business practices oriented (no line of business into which this model naturally fell), to engage in a relationship where there was no product (per se) nor revenue involved.

While eventually most of the collaborators took advantage of NEXRAD (they don't all host that) or other data at some point during the project, these challenges continued into the implementation phase - to a greater or lesser degree - with each collaborator. However, for one company, the difficulties persisted such that they had difficulty acquiring or deploying any data. In considering the challenges commercial entrants face in identifying uses and shaping government data to fit their operation, their journey provides a rare behind-the-scenes view that reveals dynamics that have not been considered in the analysis so far.

At this collaborator, the primary responsibility falls to their sales team. One of the employees involved explained the situation. He had been asked by a NOAA executive at a social gathering about their progress and began to look further into it. However, he explained "I'm not really sure how their sales organization are set up at other companies, but at our company, this isn't something that normally falls under a sales person's work" (PRI-28). Once another federal agency that was also a client began an open data effort, it started to click:

> And so, this is when I was like, you know, this is something we really need to be a part of. We need to figure out ourselves what this is. What's the business value? What's the business case? How do we get people to pay attention to it? Sounds easy - maybe it

> doesn't sound easy - but it's hard to try to get somebody to pay attention to something. At a seemingly small customer… (PRI-28)

We see then that, at least at this company, it became a matter of both identifying and selling projects upstream within the company – where the project itself, as a business opportunity, was not even on the radar. But, the primary interface to ideas, the collaborator calls, proved less than ideal:

> My head was kind of out of it because I'm a sales person, I have no idea what it means. Like, I know what weather is, I know what some of the data is, I know where the data comes from. I know about GOES-16, but when they start to get more technical and I'm like "Okay, this is why sales people should not be leading these efforts." Like, you really need the subject matter expertise within the company to be a part of it. (PRI-28)

When asked about whether they'd taken possession of any data, it became apparent that within such a large company, it was difficult to even find out if, or where, it was being used:

> This is such a large a company, and there's so many different groups within it. Over the years, found these different groups using weather data from NOAA's website, or not just weather data, but NOAA data for specific projects. And so, some we can talk about, some we can't…
>
> But it wasn't really up until a couple weeks ago that we really found a project that aligned with the CRADA. Because like other large companies, it's really hard to start an entire new business around a concept like this. You know, one of the things (the pure IAAS companies) have a little bit of a leg up against our company on is that they're just business models are more centered around data monetization. Whereas we are ... The whole problem is we're not in the business of hosting data for free unless we have some sort of business case around it, and get some sort of ROI (return on investment). And that's hard for someone in a position like mine to try to come up with. (PRI-28)

It was only when they happened upon a program that stretched across the entire company and connected to it that they were able to gain a view of operations where the data might be of benefit. He continued:

> One of the reasons why that area showed interest is because it spans across all of the company - all the various industries. That's the problem with some of this data because, it's being used in so many industries. So, there's not one person that can really give that broad view, and this team can. Where those other teams were like "Well, we don't know if this weather data is valuable. We've only seen it maybe valuable in one place."

> Whereas this team is like "No, we know this is valuable. We've seen it being used in this industry, this industry, and this industry." (PRI-28)

These are key insights. It is possible – this is a very large company – that government data, when used by only a few siloed "verticals" (lines of business within a company like healthcare, manufacturing, or agriculture), may not create dependence in a way that results in the full resources of an organization being brought to bear in exercising influence related to it. They may not even be aware of their level of dependence on it at an executive level. Or, as we see earlier with the potential conflict within a company between cloud computing sales goals and Big Data Project goals, there may even be conflicting incentives and agendas in different parts of a company as it relates to a government customer that dilute the desire to push for a particular outcome that would drive the data or its distribution in a direction that would seem more favorable overall.

In this situation, too, however, NOAA expertise is directly wrapped up in the value of the data. The sales person explains where the company is on an as-of-yet unveiled project:

> So, our team isn't looking to just host the data and provide it to the public without making the data more valuable, and adding machine learning and AI (artificial intelligence) to it. What they have in the works is creating this – basically - like a data sharing website. *What we're doing with NOAA now is, we're working with NOAA to figure out which data sets we should include first, and analyze, and see how people are using it* [emphasis added]. And it's not just sharing with users, but it's providing use cases to the public on how you can use this data. What things that they've seen being built with this data before…That's what we're doing. It's still a preview project, *and we're still in the early stages of working with NOAA on getting the feedback* [emphasis added]. It's not something that we want to make public yet. (PRI-28)

I asked him about what he saw valuable about NOAA's expertise and he answered with two common refrains: First, that his company's people are "technical" and know their tools, but not NOAA's data; and second, that NOAA experts can talk about "what they've seen the data being used for" (PRI-28), knowledge that appears to come mostly from NOAA's privileged position in the information supply chain.

It could be, of course, that this employee has a naïve view of and / or less access to the behind-the-scenes strategy and machinations of the company. However, supporting color commentary is added by his partner, a senior sales executive, who provides complementary insights into their experience with the project and the challenges presented by NOAA's expectations:

> Yeah, we join the collaborator call - recognizing that our role is as the account sales team - on the ground, you know, pointing the customer from a product standpoint. We are not the weather specialists. And I think, that is the trouble that we have found ourselves in with this NOAA CRADA because the data certainly has relevance. Assuming that an organization like ours would want to figure out what a Minimum Viable Product would be that they can leverage in their services, right? But we don't really ... That's not really our business model.
>
> I think part of it is that we don't ... You know, we don't have a group that focuses on weather, specifically, (but) we do have a very large research constituency here at the company. But it's that concept of that subject matter expert looking at that, all these incredible resources and data, and then pitching to the organization a cost model that would make sense around creating a Minimum Viable Product, that would be a software asset for the company. And that right there has been the big struggle. (PRI-29)

From this, we see he shares the perspective that the sales team is miscast in this role – and, for that matter, their organization as well. And, we see that - at least here - the "pitching" process continues outward from NOAA and into pitches inside the commercial organization, a story other companies in the project have also told. The situation gains some nuance as he continues:

> What we hear a lot internally when we're trying to pitch this idea ... Surely the (one of the teams focused on data and artificial intelligence) at our company would find it interesting to have access to 70,000 data sets. Surely, they would have something interesting to do with it. I think the response from them is "NOAA doesn't even know what that all is," but, secondarily "What does the government want?" Because we're so used to the government articulating precisely what it is it wants, versus "Hey, you guys be innovative with us!" and they're (the team) not necessarily at the forefront of government work.

> I think a lot of us were thrown a little by it ... It's not that CRADA, that OTAs, don't exist. OTA is another kind of research-style agreement that is outside of any kind of federal acquisition regulation, and it's called the Other Transaction Authority. And it ultimately makes it possible for federal government to work like a commercial entity in contractual arrangement with industry. So, those OTA's are interesting, but it's a similar kind of modality, which is "You want us to do what? You want us to dream up something that the government could find useful and then get it back to you fast" - is so not normal. You know, we're having to re-explain the scenario every time we talk to somebody internally. (PRI-29)

The focus here on an expectation of a commercial relationship with government where it sets the expectations of its needs from industry uncovers a dimension that may be lurking undiscovered in other interviews. In the world where government's (data) is an asset, versus something of industry's (material or services), the design of government procurement is stood on its head. Even ignoring the lack of domain knowledge by industry, or the choice of putting hosting platforms at the forefront - whose existing lines of business are not designed to make money off marketing the use of free data - the perspective this interview reveals is that the business opportunity looks much different than those normally presented by government. Government is asking a set of industry collaborators to make something valuable to themselves that they know is already valuable to some companies – something that is free, but government would like something in exchange for it based on the value industry will create. Akin to the earlier reference to "magical thinking" (PRI-23) - which questions whether the value proposition is really there - this brings in the idea that government is not really procuring anything, simply attempting to design a form of data distribution that they believe can produce value (for industry and others) that they can then trade upon. While we saw the genesis of this hypothesis by NOAA executives in Chapter 4, one can see that the logic is still foreign to industry and is apparently not one-size-fits all.

Framing the Big Data Project: What Happened?

Policy Innovation or Work Around?

While the BDP is presented by NOAA as innovative, it is simultaneously presented as operating within existing policy. However, the concept of operations for the CRADA sets in relief a more nuanced view of access to government information as a spectrum: While in theory the same for paper records in a dusty file cabinet as it is for streaming satellite data, the access methods and variation in value associated with them are instead bound up in the practical meaning of both open, and equitable, access. The policy workaround by the project manifested itself in two significant ways. First, under the label of a cooperative research and development project (an “experiment”), the government provided preferred access to data to a limited set of very large, private sector hosting companies. While this data was open and obtainable (technically) via NOAA’s normal distribution channel through NCEI to “anyone,” access to this volume of data was impractical for all but perhaps the very largest and most experienced customers who were willing to make the investment. How impractical? In the case of NEXRAD (discussed further below and at length in Chapter 6) an article co-authored in the American Meteorological Society’s journal the Bulletin of the AMS by both private and public employees involved in the project includes this detail “NCEI offers an offline order fulfillment option, but this option is still limited to 0.5 TB per day, with a cost of $753 per TB required to pull and transfer data off the tapes. *The NEXRAD Level II archive, requiring 270 TB in October 2015, would have cost $203,310 and taken 540 days to transfer to a single customer via an offline order*” [emphasis added] (Ansari et. al, 2018). Because of work already done by a NOAA partner / intermediary, the CICS-NC (see references to this organization in Chapter 3 as well), over 100TB of this data had already been reviewed and subjected to quality control. This data was transferred to several of the collaborators without charge and put into use immediately by Climate Corporation. (Ansari et. al, 2018).

The next workaround relating to equal access occurred when this group of commercial firms were granted access to NOAA experts and given the opportunity to jointly identify and explore opportunities to host financially valuable data that would be used by their commercial clients. While the CRADA includes a provision for them to broadly collaborate with others seeking access to the data through them, as a practical matter, this special access to expertise, with their deliberations protected from transparency by a formal, legal agreement, constitutes a variation – not necessarily a violation given the CRADA vehicle – of the information policies currently in place. In the case of the Climate Corporation, this gave them a channel through which to incentivize NOAA (and Amazon's) investment for no cost in the name of proving up the model. Again, this idea of access to expertise as it relates to a) knowledge of other external parties' use of the data, b) the attributes and construction of the data, or c) even bringing to public awareness that it exists at all also turns out to be part of what increases or decreases the value of what is already nominally open.

The Big Data Project's Achievements

While in a conversation late in 2018, the head of the Big Data Project stated that he felt the model was working (PRI-18), what "working" means in the context of the BDP is open to interpretation. Of course, the general goal of the project was to validate the proposed Concept of Operations – that is, that the infrastructure-as-a-service-providers could produce enough revenue by hosting NOAA data to cover the cost of doing so without charging the public or NOAA. Implied in this model was that if it proved scalable, it could subsidize or – ideally - eliminate NOAA's cost of distribution while increasing access to and use of the data, producing commercial and overall societal benefits. However, two important measures of success were missing from NOAA's CRADA model. First, the collaborators were not required by the CRADA to report statistics on the frequency of use / download of data. Equally important, they were not required to provide information, at least publicly, about whether or not they were recouping their costs or turning a profit from the pilot effort. The reported level of use by collaborator customers that stayed on the platform might have been a proxy for this, however, as the head of the project t confirmed:

> The percentage/proportion of users keeping the data on the platform is considered proprietary information by the collaborators. I can't divulge that under the CRADA…Sorry - but since it's their business, I can certainly understand why they want to withhold that information. (PRI-18a)

In this case, then, we are left with the information about the number of datasets moved out to the collaborators as a measure, and, to the degree NOAA or the collaborators share it, any additional information on usage in their public presentations.

Throughout the Big Data Project, the project team has communicated and promoted its achievements through presentations at conferences and to professional and government associations. Toward the end (Fall 2018) of the CRADA's life, a presentation about the project asserted that about 40 datasets had been moved to the collaborator platforms so far (Kearns, 2018a). While a list of these datasets had not been publicly presented, it is important to recognize that they might not all (yet) be publicly available – in fact, while the CRADA implies the data will be made public, it does not explicitly require it. And, to the degree they can be accessed by the public, their availability is not uniform across the collaborators. For instance, as of the time of this writing (fall 2018) one collaborator has yet to deploy any data publicly and, when using the public gateway to the collaborator offerings at https://ncics.org/data/noaa-big-data-project/, one is led to as many as 34 datasets in Google Earth Engine (Google, 2018), yet IBM displays only four (IBM, 2018).

Key NOAA Datasets. While 40 datasets may have been distributed, most presentations by the project team have focused on several datasets with significant use. For example, the Global Historical Climatology Network (GHCN) data had seen over 800,000 requests in its first four months of its availability (January – April 2017) with 1.2 petabytes of data being delivered representing 30 to 100 times the amount from NOAA in that time (Kearns, 2018b). The following analysis focuses, however, on what NOAA appears to view as the two most representative success stories, NEXRAD and GOES-16 satellite data. If we assume these provide the best case for the success of the Concept of Operations – and there is little hard information on the use of the other datasets available – then given that the picture is very limited, we can only draw limited conclusions.

NEXRAD Data: Releasing Pent-up Demand. In an earlier section, we learned about the history of how NEXRAD data came to be the first significant dataset shared with the collaborators. And, we also heard about the role Climate Corporation played in subsidizing that investment by Amazon, an important dimension of understanding the ecosystem of dependence. The principal focus in presentations about the project, however, is on the change produced in external access to data that came about from the collaborator hosting. This is important to NOAA as it represents both a sign of short-term adoption and of long-term potential for the project, at least if that were extrapolated from the results. In addition, in this early example, some figures were made public about the use of data in the cloud versus what was downloaded for processing outside the collaborator environment in an article co-authored by both public and private sector participants in the Bulletin of the American Meteorological Society (Ansari et. al, 2018).

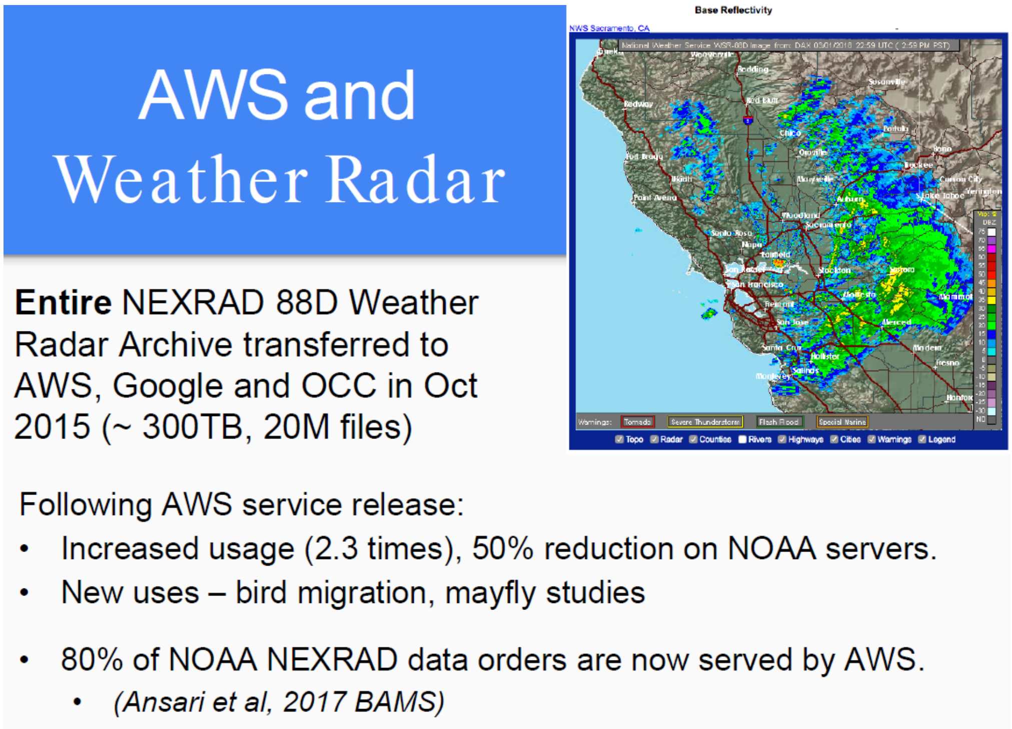


Figure 13 - AWS and Weather Radar (Kearns, 2018a)

For NEXRAD data, statistics released in the first 6-7 months of the project showed that usage had increased 2.3 times, while access obtained directly from NOAA had dropped 50% - and as much as 84% for federal and military users (Ansari et. al, 2018). In considering these figures, it should also be noted that the NCEI website through which the data had traditionally been ordered added new functionality to allow users to go directly to the collaborators to obtain the data (versus ordering it from NOAA). They also added this direct access as an option in a weather and climate toolkit they offered, creating new avenues to drive usage to the collaborator copies (Ansari et. al, 2018). Thus, one would conjecture from increased use that dependence on the data in some form is increasing and that this dependence is moving from NOAA as a distribution platform to the collaborators. In a rare case of collaborator transparency about the use of the data, usage information for this initial period was also presented that showed that 64% (221 TB) of the data accessed on Amazon was transferred to somewhere else on the AWS platform (assumedly for processing) (Ansari, et. al, 2018). One might wonder what percentage of this was solely attributable to Climate Corporation, based on the earlier discussion of their use.

The Value of Expertise: GOES-16 Satellite Data. The GOES-16 (Geostationary Operational Environmental Satellite), at the time NOAA's newest weather satellite, collects data that is highly visible – and valuable – to industry. Both interviews and highlights from presentations on the project confirm something that NOAA has begun to state more directly as time goes on, the tight coupling of underlying expertise with the commercial value of their data. In fact, the following statement was included in a presentation from April 2018 under the heading "Lessons Learned": "The key asset that NOAA brings to the BDP is not its data, but its expertise to understand and support the data" (Casey et. al, 2018). If this is true, then perhaps a growing dependence on data instead may bring with it a growing dependence on people who are wrapped up in providing its context and meaning. A brief overview of NOAA's recent experience with making GOES-16 data available via the BPD builds on this idea.

GOES-16 is the latest in a series of weather satellites to be deployed by NOAA. Up to this point, the data was only available for download via a satellite distribution network (NOAA-NASA, n.d.). Remote sensing of "earth observations" from space involves a very large amount of data – in the case of GOES-16, as much as 1.75 TB a day (NOAA, n.d.). When a satellite is first deployed, it goes through verification procedures to validate that it is working properly, and this was no different with GOES-16. The data was deployed in provisional status to the collaborators in July, 2017 (Kearns, 2018b). Once the data was officially approved, use began to rise until 8-10 times the stored volume was being accessed (PRI-18b).

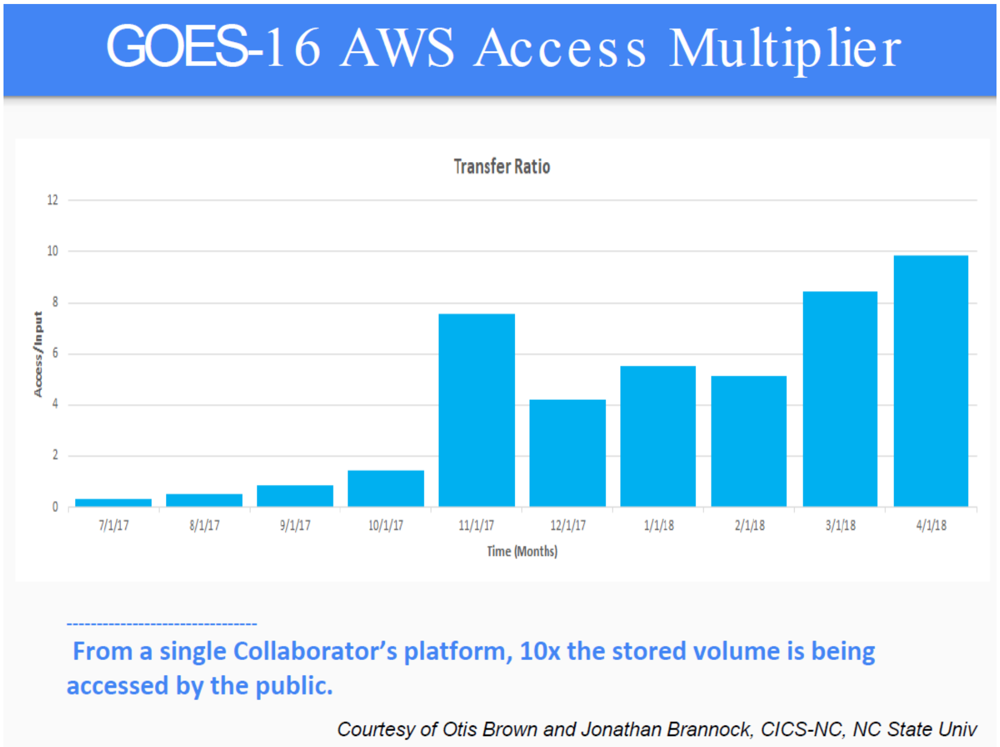


Figure 14 - GOES-16 AWS Access Multiplier (Kearns, 2018b)

Then, something unexpected happened:

> The phone started ringing at customer service back at NESDIS (National Environmental Satellite, Data, and Information Service) a lot more frequently than before. We're getting calls from all over the world from people that are interested in this data. They want to know what it is. And the customer service desk was like "We don't know who these people are." And, basically - please make it stop. (PRI-18b)

As a result, NOAA made changes to the information available on the website to point inquiries to the BDP project team who had a better view of where issues might be as well as where questions were appropriately directed. While this is a small episode, it is revealing in that it demonstrates that circles of use and dependence may stretch out quickly to new, unknown users and uses, potentially far beyond what government staffing is designed to handle, or, perhaps, policy may anticipate. If the expertise related to the data is very limited, or a higher level of expertise is required to understand and use the data, meeting this need could reach the point where it negatively impacted agency operations, or forced an adjustment in priorities for research use in parts of an agency. Thus, similar to the potential burdens on NOAA infrastructure from attempting to supply real-time access to data discussed in Chapter 3, to the degree that "human infrastructure" in the form of expertise is tied up in the value and use of the data, these constraints are also very real.

In a larger sense, however, this provides another insight that may color the question of data dependence. To the degree that the value of data – at least, certain types of data – is tied to expertise and knowledge about how it was made," what it represents, and appropriate uses (assumptions, inferences, associations), then the exercise of influence becomes more than just lobbying for particular data to be collected or organized in a certain way, or for access to be granted, or, assumedly, influencing the people directly involved in its creation and management to shape their decisions and practices. In the case of scientific data, perhaps as evidenced by the "pitch" process used in the BDP, it may be that the scientific experts sit in the role of interpreter of the data – including its potential commercial uses and value – and thus create even more reliance on persuading (or incentivizing) them of the importance of access, or specific changes in information design and collection in order to successfully exercise influence. Restated, the dependence of industry may move from the data alone to the people that create it. It is one thing for lobbyists to press for access, but this assumes some external knowledge of what would be valuable. Perhaps it is mostly insiders who can be the guide to that and, thus, have greater control over the ways it may be shaped.

The next dimension of the close bond between data and expertise became visible when NOAA took another step to expand usage of the satellite data. In the words of the NOAA project director:

> We're basically parking the data out there. And maybe we'll put out a Jupyter notebook (essentially a package of computer code, instructional documents, videos and other tools that accelerate a technician's ability to make use of data) like OCC (Open Cloud Consortium, one of the collaborators) has done, that's very successful in helping to bring

> more users in, including journalists who are using it. When I saw that journalists were picking up and using GOES data - amazing. These guys are not meteorologists, but they're able to pick up these tools and use them. We're getting eight-to-one or 10-to-one increase in usage by putting the data out in that way. (PRI-18b)

Now, in an effort to create usage – at first, mostly working with Google – he said they believe they may have found a way to multiply expertise.

> (We went) one extra step and took the data out of the scientific data format that it came in and stuffed it into existing tools, the two tools that Google's been using. One is called Big Query and the other is Earth Engine. Google has been spending their labor, because they understand these tools, and talking with our experts, then they figure out how to break down our data. Google labor puts it inside their tools, they check back with the NOAA labor to say "Can you make sure we loaded this data right?" We check and make sure they've loaded the data right, and then they turn it over to their customers. We've seen, like, at least 100-to-one increase in the usage of this data by taking this approach. That's what I call "the expertise multiplier." By having the experts, the Google expert work with the NOAA expert in integrating the data into a tool that people are already using, the learning curve about how to access the data goes away. And people just are able to look at it and go "Oh look, here's temperature data, here's cloud data, oh, here's whatever" and they start to analyze it, or combine it and join it with their data, they've got it in Big Query or the right code in Earth Engine to be able to do the analysis the way they want to do the analysis for whatever region of the world they want to look at. (PRI-18b)

It appears the agency has learned a lesson brought up previously by the provider of a private sector marketplace mentioned earlier who found that "They don't want to learn the details of the data because they have their own data that they have to learn the details for" (PRI-27) recognizing that, for private parties to really make significant re-use of their data, it needs to be made available in ways and through tools they are familiar with. In this case, the scale and complexity of the data requires – again – expertise from NOAA to make that happen. The part that is remarkable is that for this model to scale, we are again seeing that NOAA is having to provide more resources to vendors to enable them to successfully profit from hosting the data, and in turn defray the cost of free access / hosting. The cost is increased somewhat – that is, the opportunity cost of the internal, mission-related work these experts are normally assigned to – but with an increase of 100-1 in usage, it may make the project feasible for the collaborators and hopefully would not be a recurring investment after each data type is loaded and configured.

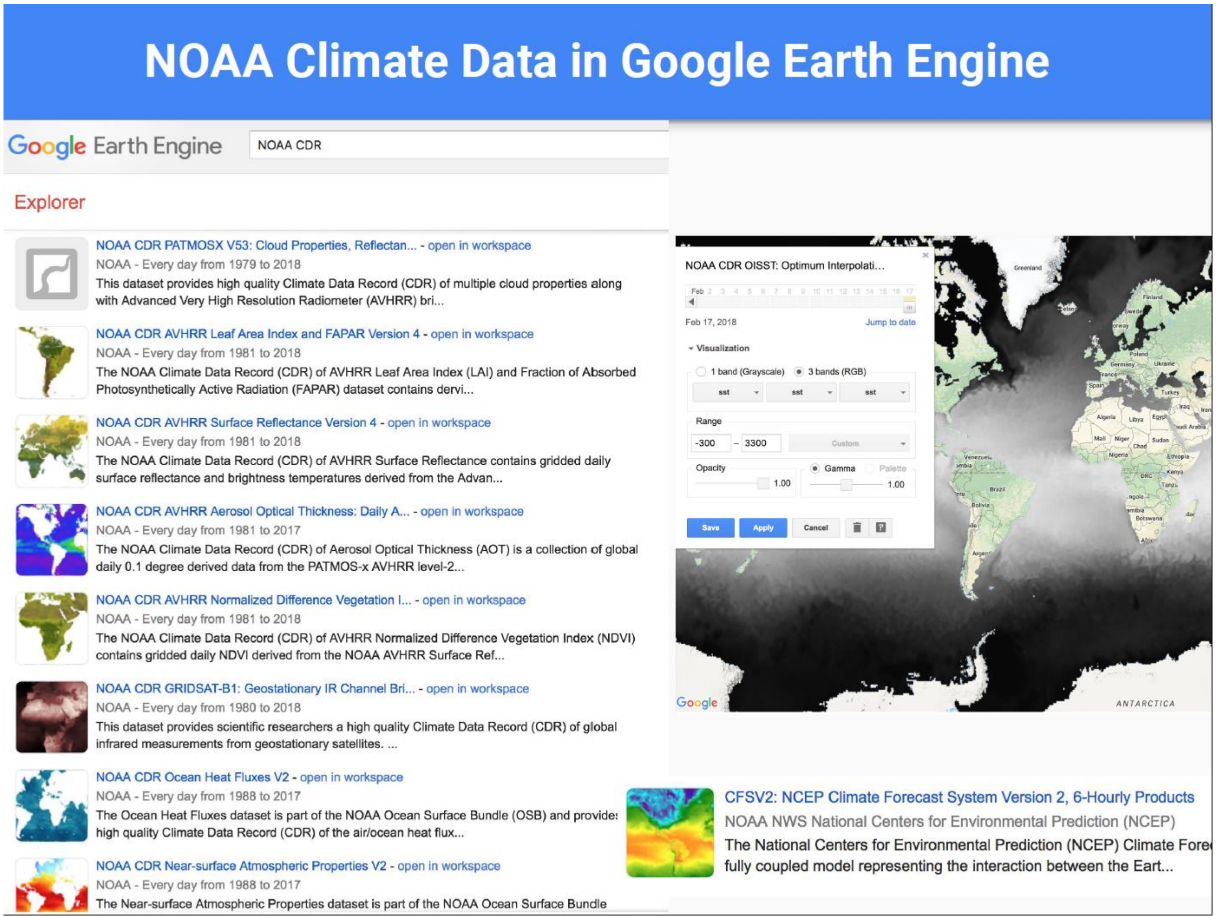


Figure 15 - NOAA Climate Data in Google Earth Engine (Kearns, 2018b)

The question for this research is, of course, how or if this volume or reuse might start to shape how the data is collected or designed – its format is already being changed. From the project director's point of view, this challenge is addressed only in passing, seeing the external and internal needs as separate, and separately motivated:

> [W]e're having a lot of discussions with Amazon and Google now about the best way to change formats and do these translations in a partnership. Because, of course, the user base wants all sorts of different formats and their needs are changing all the time. For NOAA, we're using these data for our mission purpose. There's a reason why we chose these data formats. We've optimized our systems for them, we're not going to change those anytime soon. (PRI-18b)

So, it appears that, even with strong commercial demand, there is path dependence embedded in NOAA computer systems in the form of sunk costs based on technical choices about standards that would take significant effort and financial resources to unwind over time.

Doubling Down on Expertise: The Grand Trade

However, this is not the end of the story, nor the insights with regard to expertise. In the final section of this chapter, we will see that, as the CRADA nears its end, NOAA issues both a set of brief Lessons Learned and yet another RFI to explore a more permanent relationship with a vendor or vendors to further the project. However, something telling is said by the official in charge of this effort that explains an important dimension of how he has come to view the relationship being established here, a dimension that hinges on the value of NOAA expertise:

> So, within NOAA when people say "Is this going to work?" I say "Look – it's already working." These companies already recognize there's enough value in public data sets that they're willing to have this open call for datasets of any sort (referring to Google and Amazon's recent initiatives – outside the BDP – to offer to host a limited set of open government data free of charge). And with this special relationship with NOAA that we're going to be competing, they realize even more value*, because they're going to have access to our experts in a way that nobody else will, because they're willing to make a trade for access to our experts* [emphasis added] for a service to the taxpayer of making our data available in this full and open manner. Because, otherwise, they could still have the data if they wanted to, but what they're signing up for as you know is not to sell the data but to sell these other services around it. But the data that sells are still freed up. And that's a really important concept, and that's what I've been calling the "grand trade." (PRI-18b)

The grand trade, as it is conceptualized here, now sounds more about access to NOAA expertise to help companies find a market and use for data traded for free hosting space to allow NOAA to avoid growing costs of distribution. This is interesting for another reason not touched on so far. It appears that it is not unusual for the private sector companies that use and exploit NOAA data for commercial purposes to have been founded and / or staffed by former NOAA / National Weather Service employees who make use of that expertise to commercialize the data. Besides calling into question how long NOAA staff would assist in commercializing data before they left, or were hired away, for much greater financial benefit, it actually serves to reinforce this perspective on how important this expertise is to commercialization – in many cases, at least with weather data – it appears to be a significant part of the foundation of that industry.

NOAA's Lessons Learned from the Big Data Project and Another Request for Information

Big Data Project Lessons Learned

With the end of the CRADA coming in May of 2019, NOAA released a "Request for Information (RFI) for Cloud and Data Access Services" on October 1, 2018. The RFI was accompanied by a two-page document titled "BDP Interim Lessons Learned - Executive Summary" that includes nine of the "top lessons discovered" in the project (NOAA, 2018a). Like when the project began, the new RFI is promoted as a precursor to a competitively bid Request for Proposal (RFP). As we close this chapter, the Lessons Learned and NOAA's new RFI can be helpful in understanding at least the public face of the knowledge gained by the project, and, like the previous RFI, the questions asked / statements made can provide insight into what they believe is important moving forward.

Of the nine lessons, at least several are directly relevant to our research questions:

- **NOAA must partially subsidize cost – and recognizes the value of their expertise.**

  > 4. Sustainable Public Access to NOAA data via Public-Private Partnership(s) is possible with some combination of (1) a fixed cost to NOAA, (2) cost recovery strategies based on services built upon NOAA expertise and data, and/or (3) innovative recovery of the marginal cost of data distribution from the users of the data. (NOAA, 2018a)

  This is considerably different than where things started. Here NOAA acknowledges that they may be required to provide funding for the concept to work. This is, of course, to the favor of industry, reducing their subsidy. The notion of cost recovery strategies appears to remain the same – this appears to be a euphemism for both selling computing cycles on the collaborator platform and other services. But, the foundation is now explicitly acknowledged to be built upon data and NOAA expertise – a limited resource that comes at an opportunity cost, if not an actual cost, to NOAA. And, finally, *the innovative recovery of the marginal cost of data distribution from the users* is essentially handing off this principle under which NOAA operates – its ability in policy to recover marginal costs of distribution – to the private sector to find ways to accomplish it innovatively, hearkening back to the observation by one of the collaborators about NOAA's "magical thinking" above – it is not clear what this innovation might consist of. What is also not clear here is what the ratio of each of these elements will end up looking like. Conceivably, for example, a straight payment-for-data hosting model could meet the criteria of a combination – industry seems unlikely to reveal what revenue is being produced from their cost recovery methods, so the allocation here will be difficult to discern – and challenge.

- **Acknowledging the value of a different type of intermediary.**

> *5. The role of an intermediate "data broker" has emerged as a valuable function that enables the coordinated publishing of NOAA data from federal systems to Collaborators' cloud platforms, and could possibly become a Common Service that could support all of NOAA in the publishing of data to the cloud.* (NOAA, 2018a)

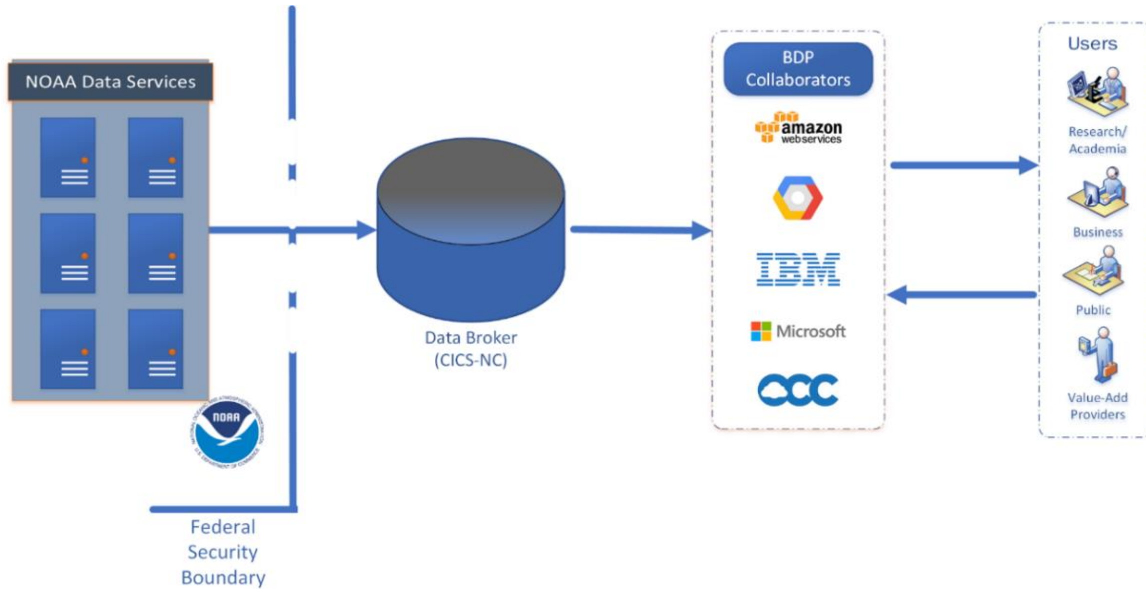


Figure 16 - Big Data Partnership Model (NOAA, 2018a)

Figure showing data distribution through the current Big Data Project partnership model with one-way transfer out of federal systems. This model allows for only a trusted user inside the security boundary, distributing a single copy of data that can support all users.

This lesson acknowledges the role played by CICS-NC. Again, while the Cooperative Institute for Climate and Satellites is an "intermediary" and trusted partner, they are a quasi-government entity under contract to NOAA. Their role does appear critical for high-volume or significantly complex data as they manage the "hop" from internal NOAA networks to the outside, acting as an organizational and technical "firewall" for security. And, in both NOAA's published document and several conversations, the advantage of this relationship as it relates to NOAA's technical security is stressed. How is this relevant? It may mean that, because of the strategic national security aspect of NOAA's mission (weather is strategic to the economy and national defense), NOAA has a reason to limit industry's ability to drive their information supply chain (and related technology requirements) directly into NOAA's technical infrastructure While influence might be expected from the private sector even on an intermediary such as this, it is not clear yet how a) such a relationship might relay influence or its effects into NOAA, or b) how an organization like CICS-NC might exercise influence themselves. As context, one can hearken back to Chapter 3 where NCEI's marketing of data for climate adaptation is being orchestrated with CICS-NC, as well as their convening of the industry to improve "Climate Normals." In a later section, we will also see some elements of the Request for Information that hint at NOAA's view of the possibilities here. Finally, in light of this lesson learned, out thoughts should return to the earlier case of the sale of Driver History Records at the state level, as a private sector intermediary contracted to government plays a very significant role in helping government obtain benefits (and manage relationships). In that case, revenue, but here, perhaps, cost-avoidance.

- **Move from experimental to a defined commitment.**

    6. *A defined commitment and level of service has emerged as a need for both NOAA and the Collaborators for the Partnership to be sustained"* (NOAA, 2018a). This lesson appears to show that NOAA has come to recognize something that industry was already pointing this out in their questions and comments during the first RFI that initiated the project: NOAA can't build dependence on a data source / industry benefit whose availability is not guaranteed, especially when they provide other options. But now it is quantified and reemphasizes that both parties must make commitments, thus binding the relationship – and related dependencies – tighter.

- **NOAA Expertise is the most valuable asset**

    > *6. The key assets that NOAA brings to the BDP are both NOAA's environmental data and the subject matter expertise to understand and support those data. The scarce asset that NOAA brings to the BDP is the expertise in the collaborative relationship with industry, and thus the most valued.* (NOAA, 2018a)

    This is a telling self-assessment. It further doubles down on emphasizing the value of NOAA expertise in addition to data, essentially commodifying them as part of the package. Or, restated in purely commercial terms, the value of NOAA's expertise can be, if not bought, traded for with an implicit price tag attached for its contribution to revenue production by the private sector.

NOAA's New Request for Information

As the CRADA is coming to a close (in May 2019), NOAA issued another Request for Information (RFI) about "[W]hether accessing the data in this manner is beneficial, and if so, the details on these benefits" (NOAA 2018b). While it is a request, an introductory section to the questions provide a succinct view of how NOAA views the core problems they are trying to address:

> With about 70,000 public datasets registered at data.gov, NOAA has a need for an effective way to easily and readily share these data in a fair and equitable manner consistent with U.S. law. However, there are two main obstacles to doing so: 1) the current data distribution model presents costs and challenges associated with increasing data volume, cybersecurity concerns, and bandwidth limitations, and 2) there are scientific content, formats and vocabularies associated with the data that require informed interpretation that can limit the number of users…(NOAA, 2018b)

The first point, of course, echoes the early discussions about the project, but leaves unspoken NOAA's ambition to foster jobs and innovation through new or wider use of their data. And, while the second point was certainly an issue NOAA was familiar with at the start, this is the first time this is publicly stated and appears to reflect a lesson learned during the project.

The RFI poses 68 questions, broken down into three categories, responses to which will be "compiled and made publicly available" - unlike either of the earlier RFIs. The first category targets End Users (23 questions), asking about use of data available through the collaborators and its volume (whether downloaded or processed in place - see relevance discussed above); whether users would be willing to pay to download it (but it would remain open "in place" on the platform); if any jobs have been created from this activity (remember the role this played in origin story in Chapter 4); and their level of satisfaction and ideas for improvement. It also asks questions about how having commercial property rights associated with the data might affect the potential economic growth from basing products on it. Several thoughts come to mind when considering what intentions might underlie such a question – if the collaborators end up making the open data they get proprietary in any form, NOAA wants public input saying that this would have a negative impact on economic development and innovation, giving NOAA leverage in negotiation. Or, if the opposite, leaning on this feedback might allow them to support it. In turn, it is possible that, as NOAA continues to buy and use private data for use in their mission, users would likely have to pay to license it. The RFI then asks how important it is to have a choice of cloud providers and takes a final opportunity to ask users what NOAA datasets they find most valuable (NOAA, 2018b) – a core question to which the answer still seems a work-in-progress.

The next set of questions – forty-two, or almost twice as many – are directed at Cloud Providers, with four additional questions on the overall initiative. These questions are cast generically to all companies in this industry. By comparing the kinds of questions asked now to those in the original RFI recounted in the last chapter, some light can be shed on both lessons learned and NOAA's anticipation of the range of possibilities going forward. One new group of questions inquire about aspects of moving significant NOAA operations to the cloud, of which public data access is a dimension – best summed up by:

> If NOAA did their end-to-end data ingest, processing, and distribution within a vendor's cloud environment, will that vendor provide financial benefits (e.g. cost reduction/credits, etc.) for these utility services if NOAA also allows open data distribution rights to that vendor? If so, what business model would be proposed? (NOAA, 2018b)

This includes questions about combining satellite and weather model data on the same cloud infrastructure and asking vendors about high-performance computing resources they offer for "massively parallel, tightly-coupled applications" (seemingly a reference to numerical weather prediction modeling, but that is not explicitly stated); whether NOAA should go directly to a cloud provider or through a third party for cloud services for internal use, if establishing "mission" (internal) services and public-facing services should be linked under the same contract, and what security schemes would be used to make sure NOAA's internal data processed in the cloud could be shared with the public (NOAA, 2018b). For our purposes, the point here is that NOAA's published Lessons Learned indicate they recognize they may end up having to pay something for hosting and it appears here that they are considering a move toward an infrastructure approach similar to the European model - discussed in the postscript below – that would place data close to compute (processing resources) that is also publicly accessible. It is not clear what new possibilities for dependence this may produce, given that information is not available showing that reuse of the data being hosted currently by the collaborators can adequately subsidize the cost of providing free hosting and access.

A second group of questions abstracted from the RFI provide insight into the options being considered for that business model, for example asking for feedback on the concept of a third party non-profit serving as a "digital bridge" (NOAA, 2018b) to cloud providers (the role CICS-NC plays now, shown in the earlier diagram); questions about data consortiums and their feasibility – if formed by industry and if fees were collected from cloud service providers to fund them – and whether or not they should be allowed to develop Service Level Agreements with each other to provide enhanced services beyond what NOAA provides. This is abstract and a little difficult to unravel, but it seems most likely to mean that intermediaries, but maybe platforms, could provide value-adds to the data; whether cloud providers are more interested in a single or multi-cloud solution and how data might be replicated between them; and then a reference to the "grand trade" discussed earlier: "Would you be willing to provide cloud storage and access capacity in exchange for NOAA's data and expertise?" So, it appears they are trying to identify ways that other combinations of organizations or relationships might be feasible, while still looking for feedback on whether or not their part of the trade is considered valuable enough by industry to make this concept financially feasible. Because the data is free to access, this is essentially attempting to determine - as noted in the RFI provision #6 discussed earlier - the market value of NOAA expertise to cloud providers. It is interesting that NOAA seems to continue framing the question of the value of their expertise to an intermediary platform company, versus trying to understand the value to commercial firms who might directly use the data. Of course, to do so would immediately reintroduce the previous problem of privileged access," but now to expertise versus the data itself.

The final grouping of questions significant to this study addresses the financial model for making the open data available. In the end user section, there was a question about willingness to pay for download (versus use in place). Here, we see questions about several aspects of that issue including whether cloud providers would charge for the marginal cost of distribution, and, if so, how that calculation would be made; whether charges would differ by user or method of retrieving data; whether other units of the federal government would be given free or discounted access; and, if a base level of access were provided free of charge, how the cloud provider might address higher requirements of users or the government. The RFI then closes with a question that seems fitting given the recurring theme of equal access throughout this story: "[H]ow will the cloud vendor ensure all public consumers are receiving the data in a fair and equitable manner (e.g. same latency, etc.)? (NOAA, 2018b)

Like the previous requests for information, it is not clear how or if NOAA will adopt a solution informed by this information. Nor, including the Lessons Learned discussed earlier, is it obvious that they have arrived at a demonstrably workable arrangement as part of their cooperative research in the BDP. What is most striking is this seemingly new consideration – there are other cloud initiatives going on at NOAA that are not focused on Open Data – of moving internal or “mission” operations to the cloud and then connecting that to public access. One is quickly put in mind of the conjecture made earlier by one of the collaborators - that perhaps the whole project was actually just a trial run for NOAA moving to the cloud, noting this excerpt from the RFI goals: “Consider implications of possible future migration of NOAA data operations to cloud infrastructure, including environmental modeling and satellite processing operations.” The outcome of this evolution remains to be seen: As of this date (February 2019), the evaluation is apparently still underway while the BDP continues.

Postscript: European Data Distribution Alternatives – ECMWF and Copernicus DIAS

When a former high-level official at NOAA discussed their prioritization and funding allocation process, she brought up a different kind of influence on decision-making:

> [W]hat we try to do is say, "This particular observation is going to advance our forecast by this amount.” Then, if you're only making incremental advances by having this million-dollar system deployed, is that really worth a million dollars? It's a hard question to answer in regards to your return on the investment. Of course, you're going to have private sector entities out there saying, "Well, in the beginning, you're only going to have this much advancement, but as the initialization of this data set until the model matures, you're going to leap frog past the Europeans in regards to this model." It was funny because the Weather Service, and in particular, the director of the Weather Service would always get upset when there'd be an article out in the media saying that the European model was doing better than the US models. He took it very personal and wanted to put in his own op-ed saying how great the US models were. My take on it was, any press is good press, right? If they want to say that the European models are better than our models, well then Congress is going to hear that and likely give us additional funds to try to continue to beat those dang Europeans. (PR-12) – (See also Kwon, D., October 1, 2015, “Are Europeans Better Than Americans at Forecasting Storms?”)

The idea of competition from another country (or group of them) in predicting weather in the United States may not be intuitive, but it is already here – on local news, in the middle of Kansas, on the weekend before Thanksgiving travel:

Figure 17 - WIBW Weather Forecast – November 22, 2018 / GFS Model (WIBW, 2018)

Figure 18 - WIBW Weather Forecast, November 22, 2018 / Euro Model (WIBW, 2018)

> "Snowfall totals from the GFS Model at the lower end here…the Euro Model - bigger snowfall, bigger snowman – so we're looking at some spots approaching 7 inches" (WIBW, 2018). (Note the difference in the amounts forecast for Sunday at 8 pm in the two models for Marysville, Kansas, for example, in Figure 17 versus Figure 18)

Competition from European weather forecasters has not gone unnoticed by the U.S. weather industry and, in turn, neither has a new EU initiative that presents an alternative model for data distribution to that chosen by the BDP, promising greater real-time access to weather data. To discuss this alternative, a weather executive now working at one of the BDP collaborators returned to the concepts behind the aborted "Reston Project" addressed in Chapters 3 and 4 - bringing the users to the data:

> [T]his issue of the data access is not limited to NOAA. NOAA is one of the big players in the world but it's not the only player in the world doing the same thing. You've probably heard the American versus the European model vernacular? For the most part, they're talking about this model called the ECMWF (European Centre for Medium Range Forecasts), which is run by a consortium at about 20 countries or so in Europe. There's also the Brits who have almost the same thing, but they have the problem. The Japanese have a problem. The Australians have the problem...the Canadians have the problem. Everybody who runs one of these major numerical weather prediction programs has the same problem. They are going to be able to create a lot more data than they can transport.
>
> So, this issue that we've always been talking about with NOAA, we've been talking with these other major providers as well. I'm pretty sure this is public knowledge. They certainly didn't tell me it's not public knowledge, but the Europeans are starting to build their own cloud services, the ones that run the ECMWF. They see this as a problem and they're starting to build the solution. Their plan is to do effectively exactly what that proof of concept was that we had envisioned whatever it was five, seven years ago (the Reston Project – see Chapters 2 and 3). I've been working with them to try to, to help them design what these services could be. So, they have plans over the next couple of years to roll out some kind of cloud services that they own or host or operate or there's a third party that they contract to operate and do the same idea. (PRI-11)

These vignettes bring to the fore another factor in the competitive landscape that could impact the value and distribution of NOAA data: Weather data collected and distributed by the European Union. We are reminded again that NOAA is participating in an information market, with multiple sources of data available that may serve similar or identical purposes and be made available using distribution models that are different from what NOAA chooses. It appears that it is not just the data, but the choice of how access is enabled and its "fit" with industry needs that can be a factor both in the value of the data to the market and the strength of industry's dependence on it. In turn, because of NOAA's dependence on commercial use of their data, this "substitute" data, and the chance of businesses using it, have the potential to impact NOAA decision making and priorities.

The European Context

While available space limits the presentation of an in-depth history of the evolution of policy on environmental data sharing in the European Union and its member states, it is helpful to know that the commercial weather sector in the EU is significantly smaller than that of the U.S. According to two studies in 2012, the EU produced about $380 million (U.S) in revenue at that time, versus an estimated $2.5B in the United States – both figures are likely significantly larger now (Pettifer, 2015; AMS, 2012). U.S. policy makers have often pointed to the U.S. policy of providing data for free or at the marginal costs of distribution as a driving factor in this difference in commercial activity and innovation, as Europe has traditionally operated on a revenue generation or full-cost recovery basis for distribution of their data (Pettifer, 2015; Venuti, 2017). Over time, however, the EU has made it a strategic priority to both consolidate the processing and distribution of their environmental data (from space via an increasing number of satellites and via "in-situ" earth-based instruments) and, as part of a larger EU Directive on the Re-use of Public Sector Information (European Commission, 2013), offer access to data for no charge. The European Centre for Medium Range Forecasts (ECMWF) is an example of their consolidated investment in high-power Numerical Weather Prediction, the results of which are seen in the graphic of the local forecast above ("Euro model"). But, it is another effort in which the ECMWF participates, the Copernicus program (formerly Global Monitoring for Environment and Security) that has recently begun a data distribution initiative with a focus and timeline (2016-) parallel to NOAA's BDP. A brief overview of this project reinforces the idea of the discretion available to government in designing data distribution mechanisms and the potential impacts differing choices may have on government and commercial data users.

Copernicus Data and Information Access Services (DIAS). The DIAS distribution architecture makes Copernicus data (around 12TB/day) (The DIAS, n.d.) available on cloud platforms, along with processing tools, in six thematic areas: ocean, land and atmosphere monitoring, emergency response, security and climate change (Copernicus in Brief, n.d.). Data from other sources can also be uploaded and combined for processing. The DIAS services were awarded to five consortia, four of which involve for-profit companies, the last a combination of a government/non-profits, each funded by the EU in the amount of $15 million euros for up to five years – with the consortiums expected to become self-sustaining through additional paid services (Geoeconomy, 2018). The concept, much like what was described by the collaborator executive above, is designed to bring the user to the data, with Copernicus DIAS consortiums running the "back office" (data store) and the "front office" space for real-time processing against it brokered by third parties (Figure 19 below). Again, however, the model shares in common with the BDP that this data continues to be available for download via pre-existing locations if a user chooses that option. The overview slide below it (Figure 20), while somewhat cluttered, highlights both that data will still be available from the traditional sources (red arrows showing "Data for Download"), while the box in the center labeled "Data and Information Access" (DIAS) contains the key principle "Data/Information co-located with IT exploitation environment and tools" that, while similar to the BDP's goal, anticipates a large portion of all the data Copernicus produces being made available that way, including real-time from satellites and other land-based sources.

Figure 19 - DIAS Concept (Veispak, 2017)

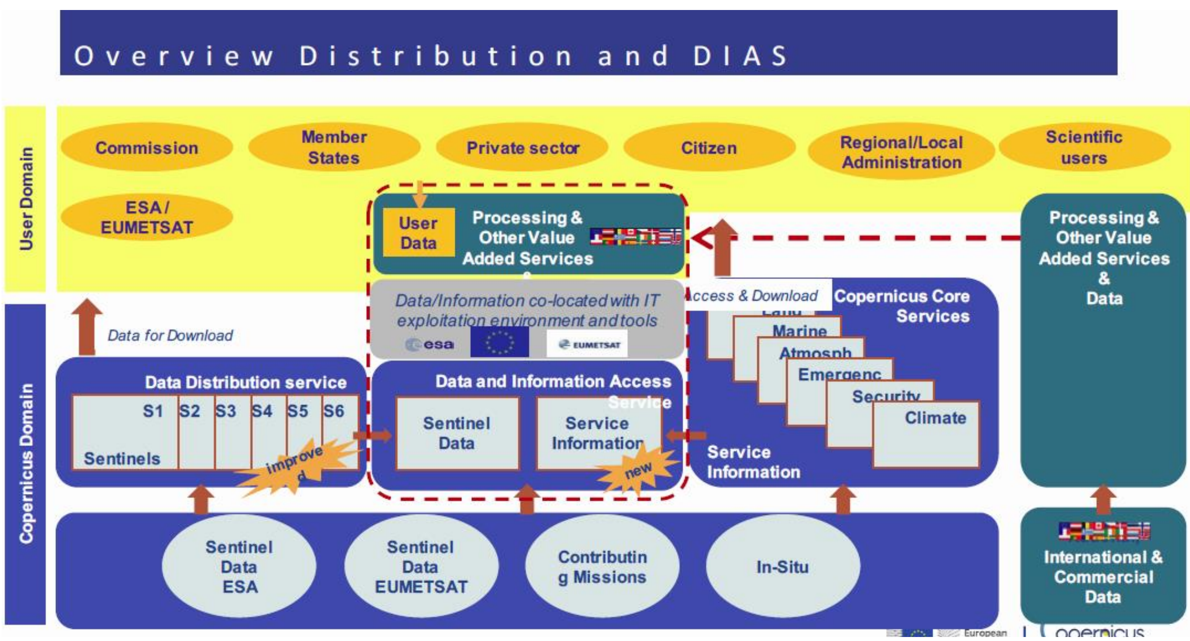


Figure 20 - Overview Distribution and DIAS (Veispak, 2017)

This model, too, anticipates that some services running in the cloud directly on the data would be free, with a higher tier of broader and more robust services available on a fee basis (Vingione, 2018).

As of January 2019, the full effort is underway, but initial results have yet to be reported. The organizational and political environments are different – it appears, for example, that an effort was made to distribute the opportunities to different member states in the EU. But, it is clear that NOAA and the large weather companies they do business with are aware of the project and the emerging distribution model.

On the surface then, we see an emerging distribution model in Europe for a broad set of weather and other environmental data. The head of data for the ECMWF confirmed the rationale driving this approach:

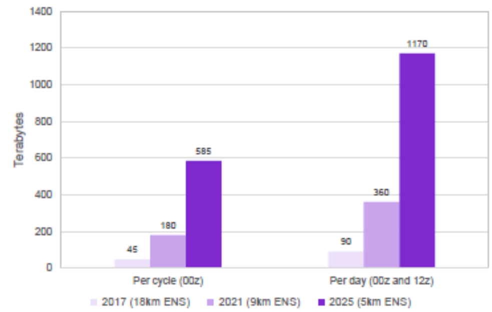
> (With) the magnitude of the archives and data in real time that we send, it's purely become more and more obvious that we won’t be able to continue a model of sending out the data in the real time to customers for very long if we keep as we are doing it, increasing resolutions of the model and keep increasing, in general, the fields, and increase the number of customers, which keeps increasing. And that will be especially true if - and it's quite possible in the future we will go to an open data model - so data will be available to the public.
>
> So, this is where we are, and we have started thinking about different models that we had a number of interactions with our users and especially at an annual meeting we have called the UEF (Using ECMWF Forecasts). So repeatedly in the last few years customers have asked us for the ability to do their own processing close to where the data produced to limit the transmission of data and movement of data. And for these reasons, we started first with DIAS and this project that is funded by the European Commission. (PRI-30)

The following two excerpts from an ECMWF American Meteorological Society (AMS) poster restate the situation:

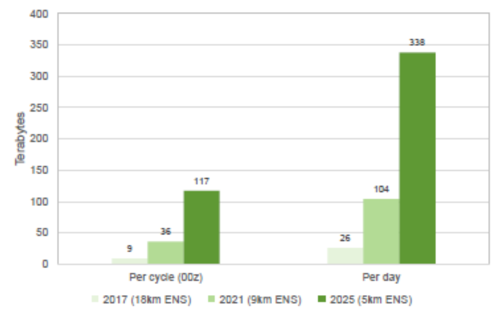
The Big Data challenge

Evolutions of the forecast models produce ever more data

- The real-time data produced will increase by a factor of 12 by 2025. Ensemble (ENS) forecasts, widely used in decision support systems, are the biggest contributors.



- Assuming current demand and delivery mechanism, the data volume distributed will increase by the same factor. Is this sustainable?



- In reality, demand is also expected to increase. The number of full data licences rose by 88% between 2015 and 2017.

Figure 21 - The Big Data Challenge (Venuti, et. al, 2018)

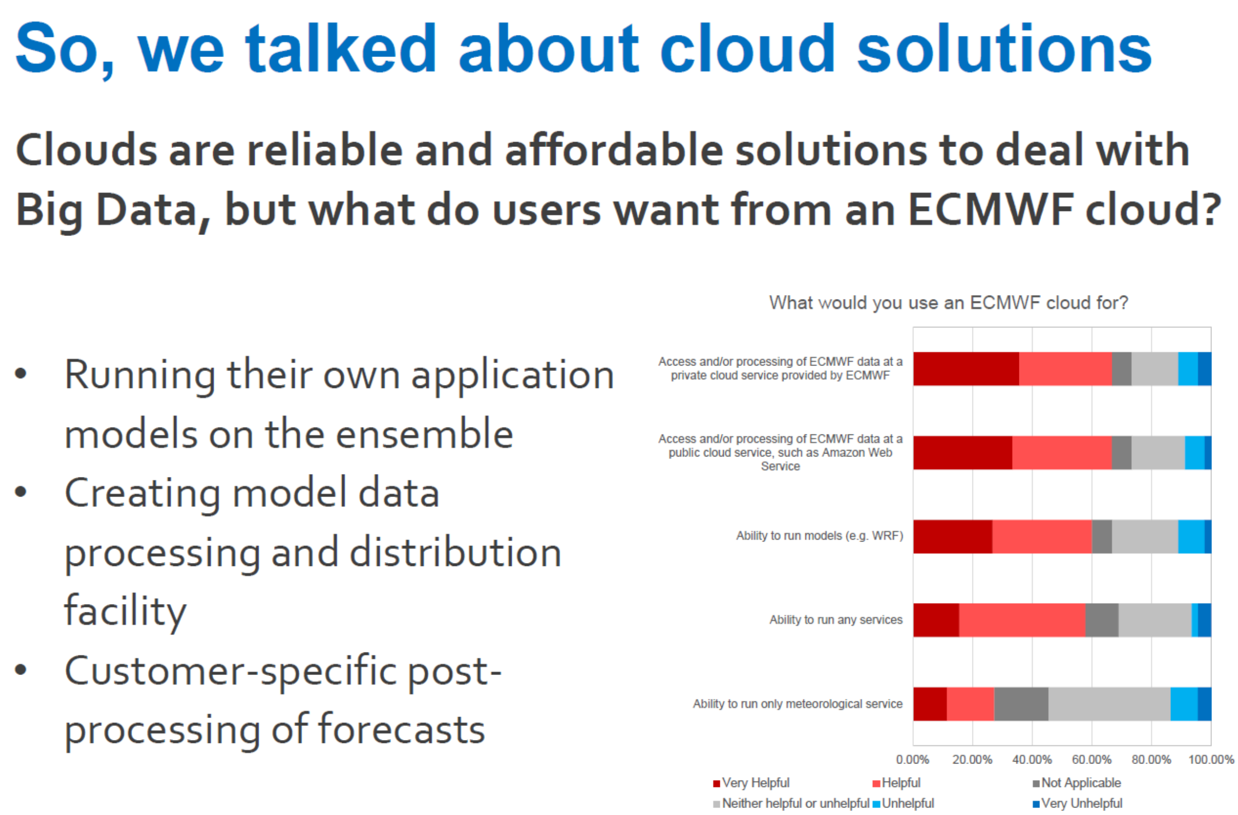


Figure 22 - So, we talked about cloud solutions (Venuti, et. al, 2018)

DIAS is an important step for the ECMWF, as it provides an environment in which to experiment to see how they will be able to meet the needs of users "with massively different requirements" (PRI-30) along with how it can be customized to optimize use of their data – the executive overseeing ECMWF data continues:

> [S]o you will go from The Weather Company who could probably digest anything - or they would like to, at least - to users, other type of customers that might be interested in doing very light processing. We are actually now in contact with a number of different users to understand better what type of user we could expect - some may be more compute intensive, but not much data intensive, if you like, and vice versa.
>
> Another aspect of the issue of exploring this area is the kind of environment that we can create in this cloud. We could have used a more standard commercial cloud solution, but the idea is that we would like really to adjust and configure it to the meteorological world, and to provide tools that are specific for processing meteorological data. (PRI-30)

The situation has elements of both technology (what is possible) and policy (what is allowed) - as he put it:

> [T]here are two aspects. One is the policy aspect, if you like, and the other is the technological aspect. The technological aspect will certainly allow users to make better use of even what is already allowable there. For instance, our ensemble data. The people that use our ensemble data use mainly surface data, and in part they use pressure level data. So, the data typically up in the atmosphere. But what is completely, totally unused is the ensemble data on model level data. So, there are either 91 or 137 levels in the ensemble. So those data are for each discrete point in the model we have 50 ensembles and we have 91 levels up in the atmosphere. These data are at the resolution of 0.2 degrees globally. So that's just the sheer amount of data is impossible to use, and then they go completely unused. So, these are data that are available already now, but it's impossible to use. It's in part the data transfer and in part is the cost of the data transfer, because we keep our network link to a decent level of operation. We charge for deliveries that go above 100 gigabytes a day, which means that forecast numbers not only may become very slow to transfer, but also become very expensive to transfer all that amount of data. So, the other question is about the policy, so despite the fact that the processing is done close to the data doesn't necessarily mean that a different type of data will be made available. (PRI-30)

Part of the calculus here is cost recovery.

> Our model, by the way, is also in financial terms economically different than the model of NOAA, because we do sell data and we do make quite a substantial part of our budget out of revenue from data sales…
>
> I think it's a very interesting time. Here in Europe we are following very closely what our member states are doing, and the way they are moving towards open data and what that actually means. What happens, we think - not actually a member state in Europe, but I think possibly elsewhere - open data is implemented with some free systems for which, however, perhaps the support is limited in time and they are not available 24/7, and so on. Then they have systems that for which support is provided and perhaps also not support only technically, but also scientific support. So, for all the user data and stuff, and this is charged…It's not necessarily value-added as in modifying the products themselves, but actually adding consultants and expertise and so on. This has been an interesting move. (PRI-30)

Here, too, we see the importance of government expertise being recognized. But, there are other dynamics here that involve the structure of the market and data sharing that are just beginning to evolve:

> The other thing that we see is, that we are following closely, is the development in the world of observations. So, the fact that private companies are entering more in this area, this market, like, for instance, sending their own satellites in orbit, or by providing in situ observations - and basically selling these observations. So how is the whole system of acquiring this data going to be affected? That's an open question. And, we are contacted - I wouldn't say all the time, but quite frequently - by companies that say that they would be ready to exchange data.
>
> But it's very tricky thing, because if we start doing more and more of these bilateral agreements, it's going to break the system of sharing global observations that has been set up by the World Meteorological Organization and that everybody has benefited from. Who knows what will happen, but it's an interesting development that may actually be very good, and may eventually end up to be the place to be, if we can have more and more observations of the climate sector. However, at the moment there is no real viable model to work with, and this is something that has been discussed pretty much at every meeting regarding data and the future of data. What we think about, for instance, the cloud services system is that if we do create an environment that can attract data providers, that is also a way perhaps to encourage this sharing of data between different actors, if you like.
>
> Because, actually sometimes we have discovered some companies do not necessarily want to sell their own observations, but were happy with providing observations as long as they could somehow collaborate with us. So perhaps even making these observations free to everybody is not such a problem for them as long as they have the ability to collaborate with us, work with us, in the maybe development of new products based on the observations that they provide. So maybe they have this edge because they have already worked with us so they have this knowledge, and they can use that to their advantage rather than merely selling of data. (PRI-30)

From anecdotal observations from an executive at the center of the European weather information market, we can see that the European Union is very focused on delivery of earth observations – more specifically weather observations – to parties external to government. And, that there is increasing acquisition of data by parties external to government that could affect the value and mix of data used by the private sector over time and the distribution and processing architecture used by government. The move to open data also seems to portend a move toward increasing emphasis on government expertise as a precious commodity (with financial or market value), the beginning of which we see in the latter stage of the BDP model as well. While the countries and their policy environments are different, it becomes clear that the creation of policies on data access and the design of data distribution models and technical architectures occur in complex issue-laden environments that continue to evolve with technological capabilities and business sophistication. The NOAA Big Data Project is only one such experiment.

With this chapter, we conclude the analysis of the NOAA Big Data Project, although the project is ongoing. In fact, a request for proposal was finally issued on April 4, 2019 and contracts awarded to three of the companies: Amazon Web Services, Google, and Microsoft in December 2019. While these contracts were not publicly available, I filed a federal Freedom of Information Act request on January 5, 2020 for copies of them and they were delivered on February 27, 2020. They reveal that, despite the announcement saying all three vendors signed agreements, only Microsoft signed one, the other two are with partners," one of AWS and the other Google. They are open for renewal every two years through 2029. While there has been little time for examination, it appears that the contracts do not offer consideration on NOAA's side, and the vendors guarantee only that they will make a minimum of 5 petabytes (about 1024 Terabytes, with each Terabyte equaling 1024 Gigabytes) of storage available. This is a large amount of storage. But, keep in mind that the NEXRAD archive was upwards of 400 terabytes, and there is streaming satellite data to consider now, too. The contracts indicate that 70 datasets have now been moved to the platform.

Conclusion

The story of implementation provides some important insights from the inside of the Big Data Project. You may recall the question asked at the Industry Day conference in Chapter 4, where a member wanted to know "If this is a contract, what are we getting?" While we may see soon what the official answer is, this chapter has provided some insight into what both parties are “getting” and what that means for our research questions.

In considering the value and theme of this chapter, one must begin by going back to the one previous to follow the expectations of both sides. Did private parties have an interest in this data? While the answer generally is yes – we have seen many companies interested in NOAA data and willing to lobby to gain access. But, in this case, NOAA has picked companies to collaborate with who even go as far to say, literally “We don’t care about your data.” By this, the executive meant the quality of it, but it was also clear that the companies involved here weren’t organized to, nor perhaps had the incentive, to obtain the benefits NOAA saw for them. Thus, like the promotion efforts at the NCEI, while industry has shown some interest, for the players involved, this is no Driver History Data situation.

The answer to the next question is informed by this disinterest, but there are some differences. When we ask how the companies involved seek to gain influence on the data or access to it, we did see early on that companies were approaching NOAA via the collaborators looking for exclusive access, and that it was necessary later, to help a collaborator increase use of the data on their platform, NOAA experts did assist them in loading and transforming it to be more usable within the proprietary tools within which it was delivered. As the BDP project manager noted “We've seen, like, at least 100-to-one increase in the usage of this data by taking this approach,” this is an area to pay attention to – as they definitely have – in that these NOAA-aided transformations may grow. Although it might not be readily apparent, we saw some influence, in this case by government on industry, as b) during the promotion and negotiation phase of that agreement, NOAA was quite willing to play the market by creating a sense of competition between them to participate, and b) they picked a set of companies, some of which did not even respond to their call for information, to enter into a confidential agreement with to cultivate ideas for more use, and more lucrative use, of NOAA data, both things that show how government was willing to exercise some influence in the service of their goals. And, industry, having more irons in the fire than just this project, was incentivized to participate to maintain good relations. One might look back to the press release and symbolic value of government's participation in the Restaurant Inspection case in this study to see that goodwill, and good relations work both ways.

In seeking to address the final, fundamental question about the impact this influence had on government, we see, I think, mostly, that it was government having an impact on itself in the hope of achieving benefits. It appeared to summarily ignore – in the previous chapter – opportunities to partner with dozens of companies in the name of its own theory of what would make the market work. And, yet, as they went forward, the most significant contribution to be derived from what happened in the study of dependence on secondary use of government data were the numerous and varied barriers encountered by both sides in establishing a model where both sides gained significant benefits and supported the feedback loop necessary to sustain it.

First, and maybe most obvious, was the continued struggle to identify data that was available and that had revenue potential. Because confidentiality agreements were in place, it is hard to delve deeper to say exactly why that might have been. But, in the words of some of the participants themselves, a key reason might be that they picked the wrong companies, who weren't organized to take advantage of the opportunity – if there was one. They were also limited by policy – it would have been seemingly impossible, for example, for them to award a contract solely to The Weather Company or a few other vendors, again, for data they likely already have access to, or new data that, again, could be shut down by others that framed it as "privileged access." One result this appears to have had is that the data so far seems mostly something that companies had access to anyway, although maybe now in a new or improved way, and more datasets are likely to be added over time.

But, there were many others. We see that government still lacks the resources internally to operationalize this – no full-time team or organizational structure is in place in addition to what existed previously. In discussions, there is an emphasis on being judicious with the use of NOAA expert's time and not disrupting operations. And, when use of one satellite dataset catches fire, it takes an emergency effort to put out the flames. So, there is a question of how well this will scale, especially if more expertise is part of the "grand trade" in the future. We also see the potential for data substitutes introduced, like with driver history records in an earlier case, in the form of European and private sector data. NOAA may not always be the sole source of data, nor is there any guarantee that, hidden somewhere in the largest volume would turn out to reside the most valuable types of data. On the private side we see barriers as well. It is not that there are not companies – it seems the weather industry would be full of them – that can find ways to add value to data that is freely available to everyone. But, certainly in the case of the collaborators, it was not their business model, and it seems like that - across the 66,000+ types of data NOAA claims to hold – much of the potential market for adding value to that data may not have a business model oriented to do that either. In both this case and the earlier one cataloging the NCEI efforts, it appears that NOAA has taken on a task akin to the oft-questioned "nation building" in the political world, where the road ahead is long and, subject to nativist tendencies, factions, and events on the ground, producing results that are at best uneven. There is another barrier, too, for business that may have been unexpected and of which this is a part. While their business models seemed entrenched in a path dependence reinforced by success, so there is also path dependence in their technical models. The idea that "not all clouds are created equal" is an interesting one for what it says about dependencies business may have on technologies, or lack of them, that make turning their ship difficult. And, if technology is defined more broadly as a facilitator of processes in the modern organization, it may be a larger barrier still. Witness the company that was unable to access information internally to understand how it might even have been using external sources of data, or, as we see in the next chapter, itself has systems based on standards and interrelationships, strengths and weaknesses just like NOAA, that make it difficult and even unwise to invest in changing until they are certain it leads to a "can't miss" opportunity. What if their internal technical infrastructure looked just like NOAA's diagram, plagued by similar challenges, yet each competitor somewhat different in what these were? Equity is a worthy ideal, but it has never been a guiding force in the market, and it appears these efforts that use government data as a resource will always sit in tension with that.

Chapter 6: Influence and NOAA Weather Radar: An Information Lifecycle Analysis

Introduction: Looking for Influence in Next Generation Radar

As we come to the last case in this study, the analysis so far has seen commercial attempts to influence government data and policy - and government attempts to influence industry - produced mixed results. When both government and industry benefited, whether it be through genealogical or driver history records, we saw signs of impacts on government policy and practice, and perhaps the potential for more. And, when government saw benefits to be gained in program sponsorship, mission objectives, or cost avoidance (or all three) by engaging the private sector (the NCEI data promotion and the Big Data Project), we saw them willing to change their practices in the hope of gaining more private sector dependence in support of those goals. Then, in the case of NOAA's policy on provision of environmental data to the private sector, even though NOAA wasn't excited about it – they even mark up industry's own conceptual diagram in their response to the SAB recommendations to show the difficulty - the sheer force of private sector lobbying was able to embed concessions in government policy. Yet, industry still struggles to obtain the related benefits in practice. And, if we return to the very first case in this study, a company wanted to get government to implement a new standard in return for delivery of its inspection results to a much wider population. But, the benefits – and motivation – for government just weren't strong enough to overcome the barriers. So we end our investigation with one last variation along this same line, but where the benefit to industry isn't just an extra feature in one or two companies' smartphone app.

Instead, it is NEXRAD Weather Radar, the data behind that familiar animation of storm fronts one sees on the nightly news, or a weather app on a phone, as they move closer and closer overhead, bringing everything from light rain to millions of dollars in destruction. The case for this choice is compelling. First, we have already seen that there is high commercial demand for it. Its benefit to society is supported by the government's own assessment, where it is ranked as the second most beneficial "earth observation system" nationally (NSTC, 2014). The societal benefits used for that assessment, as discussed in Chapter 3, come from directly from NOAA's mission and, thus, along with the size of the commercial weather industry, signal that its secondary use may help NOAA avoid costs by working through the private sector to achieve its mission, potentially fostering dependence on that use – or is that a strong enough incentive? Another appealing aspect is that, in one form or another, data from weather radars have been produced for decades, and significant documentation exists of its origin and history, including changes in its production, use, and distribution – all available to help in the analysis. Finally, considering the variation in technical complexity of NOAA's 70,000+ datasets, weather radar is a familiar feature of life for most Americans so the subject will be approachable, even if technically daunting at times.

To start this journey, we have some context. The last three cases have all involved NOAA. But this is the first to take the long view, recounting highlights in the evolution of a program and the data it produces. Taking this approach also allows us to look at the role that technology plays in dependence as it changes, enabling government <u>and</u> industry to do new things, and creating new demand to support them. In tandem, we will also move forward using an "information lifecycle" perspective that illustrates how, really with all data, each stage from its collection or creation, to its retention, processing, and delivery is a possible site of influence.

There is yet a more significant reason supporting its selection, however. Using NEXRAD as our subject creates an opportunity to consider its role in the Big Data Project as (to quote the previous chapter) "…the most recent step in a long history of interplay between content, structure, technology, access policy, distribution methods and commercial influence" that surround it. In the last chapter, for example, we heard a firsthand about the process that led NOAA and the CRADA collaborators to follow the lead of a Presidential Innovation Fellow (PIF) to deploy NEXRAD as the first dataset in the Big Data Project. Some of his stated reasoning, however, came from experience with past distribution schemes for the data, choices we will see described in the following pages. In turn, a senior executive at Climate Corporation explained the company's interest in subsidizing the transfer of the NEXRAD data to the cloud as a result of difficulties in accessing it. In both cases, the new path being explored for distributing weather radar data was informed by the past – in this case, the past of NEXRAD specifically. The new methods for distribution and processing prototyped in the Big Data Project might even be viewed as just another evolution in the paradigm of access and use – now leaving data in place on private sector servers so they can more easily make use of high-power computing to process it.

So, does NEXRAD's past have lessons for us in the present? As we follow the history of weather radar data, we focus on dependence in the early days of its development and distribution – before the enormous processing and storage capacity of cloud computing, or fairly recent federal policies requiring machine-readable data were put in place (EOP, 2013) - in search of commonalities that might answer that question.

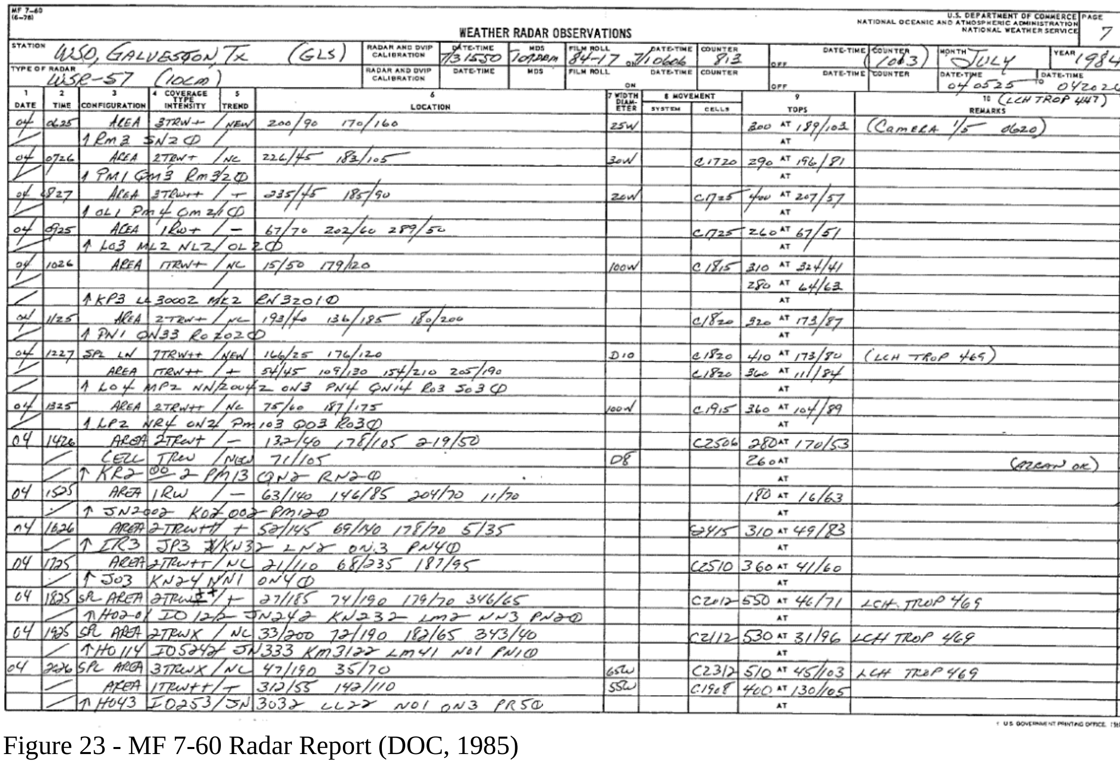
Weather Radar – Maturation and Early Expansion of Dissemination

In previous chapters we saw that difficulty in accessing data already known by industry to have commercial value was a key incentive driving their interest and influence. In the case of weather radar data, we will see that as the radar technology matured, so did computing and networking technology that allowed it to be distributed more easily both inside and outside

NOAA. This, in turn, allowed for expanded secondary use(s) – and, thus, private sector dependence. And, to the degree that data is retained (archived) after initial use, increased access can also unlock latent value for longitudinal research across a time-series of observations. On their own, increases in technological capability can lower access costs and enable more use. So, advances in technology are part of the dependence – and influence - equation. In this case we will see that it is this interrelationship between changing technological capabilities, collection, retention, and the design of distribution - and the self-interest of both government and industry - that forms the foundation for secondary use. To see those forces in action, as well as provide background for further analysis of influence and dependence related to radar data, we begin by looking at the evolution of this data and its distribution in its early days.

The Birth of Weather Radar Data

Weather radar technology traces its origin to Britain during WWII where it was used for detecting enemy aircraft and to understand and predict impacts of weather on military operations (Whiton et. al, 1998). Rudimentary compared to today's technology, the U.S. military brought these radar units back with them after the war. They were not networked and used different versions of the technology available at that time. And, they were not yet deployed to cover the territory of the United States. Meteorologists used a weather radar report (RAREP) – an early form of radar data – to communicate what they were seeing by teletype. The following examples are of the RAREP (MF 7-60), a later generation of the report still in use in the mid-1980's, along with its instructions (DOC, 1985):



MF 7-60 (6-78)

U.S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE

PAGE 7

WEATHER RADAR OBSERVATIONS

STATION: WSO, GALVESTON, TX (GLS)

TYPE OF RADAR: WSR-57 (LOCAL)

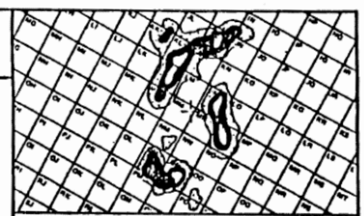
MONTH: JULY YEAR: 1984

1 DATE	2 TIME	3 CONFIGURATION	4 COVERAGE TYPE INTENSITY	5 TREND	6 LOCATION	7 WIDTH DIAMETER	8 MOVEMENT SYSTEM	CELLS	9 TOPS	10 REMARKS	11 INT.
04	0625	AREA	3TRW+	NEW	200/90 170/160	25W			200 AT 189/103	(CAMERA 1/5 0620)	

Figure 23 - MF 7-60 Radar Report (DOC, 1985)

CONDENSED EXPLANATION OF THE RADAR REPORTING CODE

LOCATION IDENTIFIER	TIME	ECHO CONFIGURATION	AREAL COVERAGE	PRECIPITATION TYPE	INTENSITY	INTENSITY TREND	ECHO LOCATION	ECHO MOVEMENT	ECHO TOP
EEW	1735	AREA	5	TRW	+	/+	8/80 182/80 90W	C2325	MT 300 AT 191/55



DIGITAL LOCATIONS AND INTENSITIES

IM22 JK2232 KK2321 LL3132 ML2023 NM1 OM231 PM 221

DECODED REPORT

At 1735 GMT the Neenah Wisconsin radar shows an area five tenths covered by thunderstorms. These thunderstorms are intensifying and some of them have reached heavy intensity. The area extends from 80 n.mi. at 8° to 80 n.mi. at 182°. The area is 90 n.mi. wide. Individual thunderstorms are moving through the area at 25 knots. The highest echo top is 30,000 feet. It is located 55 n.mi. from Neenah at 191°. The digital section of the observation shows the maximum intensity of the echoes observed in the grid boxes in rows I through P.

TIME OF OBSERVATION

The time of observation is reported in Greenwich Mean Time (GMT). Regular observations are taken at 35 minutes past the hour. Special observations may be taken at any time.

AREAL COVERAGE

The areal coverage in tenths is reported with a one or two digit number.

ECHO CONFIGURATION

CONFIGURATION	DEFINITION	CONTRACTION
Isolated echo	Independent convective echo	CELL
Area of echoes	A group of related echoes	AREA
Line of echoes	A line of related convective echoes	LN
Elevated Stratiform echoes	An area of stratiform precipitation aloft	LYR
Fine Line	A narrow nonprecipitation echo associated with a meteorological discontinuity	FINE LN
Spiral band area	Curved lines of echoes associated with hurricanes, or tropical storms	SPRL BAND AREA
Hurricane Eye	The central area of a hurricane	EYE
Center	The central area of a hurricane or tropical storm. Less well organized than an eye	CNTR

PRECIPITATION TYPE SYMBOLS

R	Rain	ZR	Freezing Rain
RW	Rain Shower	ZRW	Freezing Rain Shower
L	Drizzle	ZR	Freezing Drizzle
S	Snow	SW	Snow Shower
IP	Ice Pellet	IPW	Ice Pellet Shower
A	Hail		

The letter "T" is put before any precipitation type when thunder is believed to be occurring with the precipitation.

PRECIPITATION INTENSITY SYMBOLS

-	Light	X	Intense
no symbol	Moderate	XX	Extreme
+	Strong	U	Unknown
++	Very Strong		

Intensity symbols aren't used with drizzle, snow, ice pellets, or hail.

INTENSITY TREND

TREND	SYMBOL	TREND	SYMBOL
Increasing	+	Unchanged	NC
Decreasing	-	New	NEW

ECHO LOCATIONS

1. If the echoes are in a line or rectangular area, the direction and distance are given to each of the end points. The width of the area, or line, is also reported.
2. For irregular shaped areas, the directions and distances to the salient points are reported.
3. For isolated cells and circular areas, the direction and distance to the center of the cell, or area, is reported along with the diameter.

MOVEMENT

The direction from which the echoes are moving is reported in tens of degrees. Speeds are reported in knots. A letter identifying the type of movement precedes the direction. "A" means area movement, "C" means cell, and "L" means line.

ECHO TOPS

The heights of echo tops are reported in hundreds of feet above mean sea level. The direction and distance to the top is also given.

REMARKS

Any appropriate remarks are encoded following the echo tops. Such things as radar indications of severe weather and comment which clarify the observations are included as remarks.

U.S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
NATIONAL WEATHER SERVICE
SILVER SPRING, MD. 20910

OPERATIONAL STATUS CONTRACTIONS

Operational status contractions are encoded with the remarks. The following contractions are used.

CONTRACTION	MEANING
PPINE	No echoes detected
PPIOM	Radar out of service for maintenance
PPINA	Radar observation not available
ROBEPS	Radar operating below performance standards
RHINO	RHI Scope Inoperative. Height not available
ARNO	A/R scope inoperative.

DIGITAL LOCATIONS AND INTENSITIES

The digital section of the observation is based on a grid with boxes that are about 22 n.mi. on a side. Each grid box is identified by two letters. The first identifies the row containing the box and the second the column. Rows are lettered from North to South, columns from West to East.

The digital section of the observation is made up of one or more groups. The first two characters in each group are grid box identifiers. The next character is a number which represents the maximum observed echo intensity in that box. Any following numbers represent intensities in the following grid boxes in the row. The rows are encoded from North to South and the columns from West to East. The intensities are encoded using the following numbers.

CODE NUMBER	INTENSITY
0	no echoes
1	light
2	moderate
3	strong
4	very strong
5	intense
6	extreme
8	unknown (believed severe)
9	unknown

Figure 24 - Condensed Explanation of Radar Reporting Codes (DOC, 1985)

Technology and the Emergence of Commercial Value

While today's public may think of weather radar's principal value as the prediction of rain and early warning of severe weather, the early push to expand its use came from a desire for advance warning of hurricanes – a phenomena now primarily predicted by computer models and tracked off-shore by satellite. Today it seems almost impossible to conceive of a time when hurricanes weren't known about days in advance, yet early efforts relied on the imperfect efforts of air reconnaissance (unable to track progress at night) and reports from vessels at sea that were likely trying to escape or avoid them (Whiton et. al, 1998).

To address these limitations, the Weather Bureau placed repurposed military radar units (with a range of around 200 miles (Bigler, 1981)) near the eastern and Gulf coasts. Their value in protecting life, property – and business – became obvious right away. However, while the Weather Bureau shared information about impending storms with the public, these radars weren't yet numerous, nor was their coverage broad enough to allow industries potentially impacted by the storms to take pre-emptive action to avoid their effects in a timely manner with any degree of precision. Thus, even at this early stage, industry had the incentive to become involved in the collection and use of radar data in areas where their facilities might be at risk. In 1948, for example, Dow Chemical, along with several other companies along the Gulf of Mexico, deployed their own weather radar to improve their ability to track the potential impact of hurricanes on their manufacturing plants more accurately, reducing "false positives" that resulted in unnecessary closures, saving valuable production time. This, in a refined form, is a service that commercial weather companies have evolved to provide today (Figure 25) (Smith, 2013), customizing NOAA and other data to issue more targeted "site-specific" tornado warnings for industrial facilities than the National Weather Service provides to the public.

/ How Good is SkyGuard?

A thirty-two month study of America's automobile industry documents the amazing value of SkyGuard. An unplanned shutdown is extremely expensive (i.e., half-painted cars ruined, stamping equipment must be cleaned, robotic computers reset, etc.).

Auto Manufacturer	NWS Tornado Warnings	SkyGuard Tornado Warnings	Hrs. of Production Saved
Company A	203	25	131
Company B	54	14	31
Company C	133	14	93

This represents a cumulative saving of $15 million dollars via the elimination of lost production time to the U.S. auto industry during the period of study.

Figure 25 - How Good is Skyguard? (Smith, 2013)

Collection and Distribution Expand. A turning point in deployment of radar occurred when, in reaction to five hurricanes that struck the U.S. in short order in the mid-1950's, Congress authorized a much larger investment – as many as 30 units – along with staffing and training that formed the beginning of the national radar network we know today (Whiton et. al, 1998). This broader deployment of similar weather radars between 1959 and 1964 was coupled with new technologies for data distribution. In the early days of radar, the data consisted of the RAREP's noted above, along with communication by telephone, and then teletype through what was known as the Radar Warning and Coordination (RAWARC) network (Bigler, 1981). At one point, RAREPS were made hourly by radar operations, then sent to a location where they were consolidated, with messages summarizing the data sent "by teletypewriter about 50 minutes later (and), in 1962, maps were distributed on the national facsimile circuit at three-hour intervals" (Bigler, 1981). Then, new WSR-57 weather radars were deployed that were equipped with both Polaroid and 35mm cameras (Whiton et. al, 1998) mounted over the radarscope (known as the Plan Position Indicator, or PPI) that captured critical indicators displayed on the screen. Here we have more advanced visual “data” produced that is suitable for secondary use and, indeed, in a history of the “pre-NEXRAD” weather radar era, the authors write:

> Thousands of feet of film, produced by these cameras for almost four decades, are on file at the National Climatic Data Center (NCDC) (Note: Now NCEI, the National Center for Environmental Information). The films served as a component of aircraft accident and incident investigations when it was suspected that thunderstorm-associated aviation hazards, might have been a contributing factor. The scope photography capabilities built into these early radars *led to the multi-level digital data archiving systems of today’s WSR-88D* (Note: NEXRAD) [emphasis added]. (Whiton et. al, 1998)

Then, in 1965 came the first deployment of the Radar-to-Telephone Transmission System (RATTS-65) that allowed for “near real-time telephone-line transmission of PPI data and handwritten alphanumeric annotations to nearby offices” (Bigler, 1981). In this case, a slow-scan television camera was mounted to focus on the radarscope, coupled with a mirror to allow an image with an underlying map to be written on by a meteorologist and then superimposed on the picture before being sent to a receiver in another location via telephone line (Hilton & Hoag, 1966) (see also Figure 26 below):

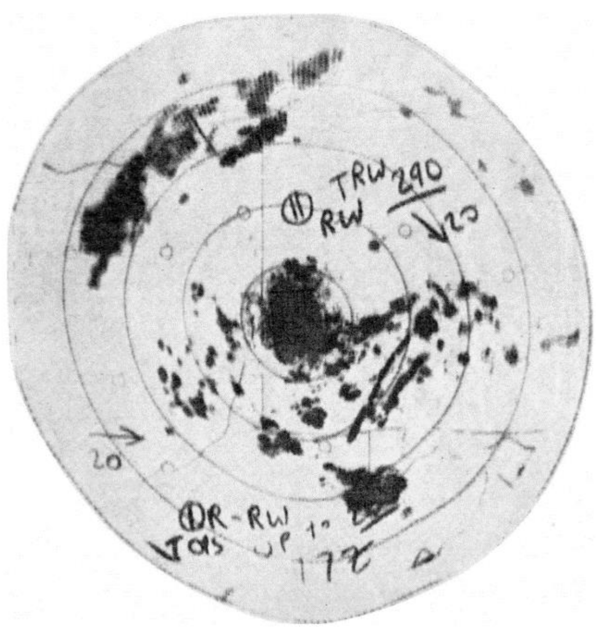


Figure 26 - Image transmitted from Pittsburg, PA to Seattle WA via RATTS-65 (Bigler, 1969)

The benefits to meteorologists in the National Weather Service from advance warnings were significant. But it was also well understood at the time that the technology opened a window to a new variety of secondary uses. An article outlining the technology published in the Bulletin of the American Meteorological Society in 1969 suggested just a few:

> The full potential of this low-cost radar remoting capability is yet to be determined. It is not difficult to propose several important uses for the remoted radar data…For example, one can envision the airline dispatcher concerned with precipitation over the company's air routes…Television stations may obtain displays of severe local storms occurring many miles away…Industrial meteorologists may monitor the location and movement of precipitation for clients in distant cities. (Bigler, 1969)

The rollout of the system would take time – into the early 1970's (Bigler, 1969) - but by 1981, an estimated 125-150 television stations were able to broadcast radar images in color. This adoption, popularizing the visualization of radar data, along with weaknesses in the RATTS approach began to shape demand even in the military:

> Shortfalls in the RATTS and WBRR systems created a niche market that firms such as Kavouras, Inc., Alden Electronics, WSI (Note: Forerunner of The Weather Company), and others eventually filled by selling services that provided timely access to attractive color, remote radar displays at an affordable price (E. Dash, 1996, personal communication in Gratz, 2005). *The wide availability of radar data on television weathercasts created a demand among the operational customers of military services for the same sort of service… (*thus*) Remote access to radar weather data was incorporated into the design of NEXRAD by making the data available to value-added resellers under the NIDS* [emphasis added]. (see NEXRAD Information Dissemination Service discussion in a subsequent section) (Whiton et. al., 1998)

This approach, then, represented the first step in enabling real-time secondary use of weather radar data. And, while these were simply pictures of the weather radar, the next step was already in sight:

> While techniques exist for automatic digital processing of weather echo data, such equipment is not expected to be available for several years. Digital data processing offers the maximum flexibility in data handling and may ultimately replace the slow-scanning technique. (Bigler, 1969)

In the background, the private sector expressed their interest in these improvements throughout this period, along with the idea of routing around NWS distribution limitations, as the following excerpt from "News from the Chapters" recounts it in a recap of a meeting of the Anchorage chapter of the American Meteorological Society (AMS, 1981):

> Finally, the results of a National Weather Association (NWA) survey, which was sent to the Communications Director of NWS, were discussed. Private sector meteorologists want a backbone high speed teletype circuit to help get information to the public, offer their own information, and relieve the NWS workload. With many chapter members belonging to both NWA (National Weather Association, a professional association) and NWS, this led to a lively discussion. It was predicted that NWS would not be in favor of the proposals.

From this brief history of radar data, we learn several things relevant to our exploration of influence. First, radar data itself has changed over time, and with more detail and faster access, its value and external demand for it has increased. Both the following analysis of NEXRAD radar data and that of the Big Data Project should be placed in this context. Next, advances in technology play a significant part in what can be done – both in observation and in distribution – and decisions about adoption shape the playing field for where influence might be applied, including the resources required to participate. Finally, we see a convergence toward standardization in observation and distribution that benefits data sharing inside NOAA and, assumedly, later efforts by industry to access and reuse the data.

Weather Service Radar – 1988 – Doppler (WSR-88D / NEXRAD)

To talk about influence and dependence in the secondary use of weather radar data, a basic high-level knowledge of how the data comes into existence, is processed, retained, and then distributed is required. The most current iteration of weather radar, and where the remainder of our discussion will focus, is the WSR-88D (Weather Service Radar - 1988 Doppler) radar known as NEXRAD (Next Generation Radar). Like its predecessors, NEXRAD sends high-frequency pulses from a fixed base station that are reflected back by varying phenomena in the atmosphere to a receiving unit. The distances at which they provide value vary from 50-250 miles based on the purposes for which they are used (Crum & Alberty, 1993). The shape and movement of these reflections are captured and analyzed to produce measurements of the atmosphere like cloud formation, height, and movement as well as the makeup of precipitation (hail, snow, rain, etc.). The base stations, of which there are now over 160 geographically distributed across the United States (including military and airport locations), are sometimes located at a distance from the nearest weather office, with the data transmitted to them via high speed connections. The radars can be controlled remotely by those offices to execute different "volume coverage patterns" like "clear air mode" or "precipitation mode" as they rotate and scan the sky. The amount of data produced in one "volume scan" by the execution of these scanning patterns both vertically and horizontally is a function of the frequency and speed of the pattern chosen, along with intensity of the weather itself.

Over time – some of the history will be recounted below – software and hardware technologies have been introduced to improve the types of data gathered - Doppler (the "D" in WSR-88D) to help detect storm movement and wind direction which was part of the original rollout, or Dual-polarization (Dual-pol) – added later in 2011-2013 (ROC, 2019) - to gain a more refined look at phenomena in the atmosphere, especially precipitation, by sending both vertically and horizontally shaped pulses as part of the scan. Yet, the basic system and data flow remains similar to when it was first implemented. With this in mind, the following diagram, taken from a journal article published in 1993, can be used to illustrate these basic relationships. Access to and distribution of the data will be addressed in a subsequent section.

The diagram (Crum & Alberty, 1993a) describes the equipment and where the data and products are produced. The tower supports the familiar "radome" (radar dome) that covers the elevated radar unit to protect it from the elements. The *Radar Data Acquisition unit (RDA),* the hardware/software that receives the data from the antenna that “catches” the returning echoes – the *Level 1 Analog Signal data*, then processes it to create the *Level 2 Digital Base Data*. This data is passed on to the *Radar Product Generator (RPG)*. The RPG – which may be collocated with or located at a distance from the RDA - creates “products” *Level 3 data* by applying algorithms to the data. Finally, the *PUP (Principal User Processor)* is the processor and workstation that uses these products to visualize atmospheric phenomena and manipulate it to create *additional products - Level 4*.

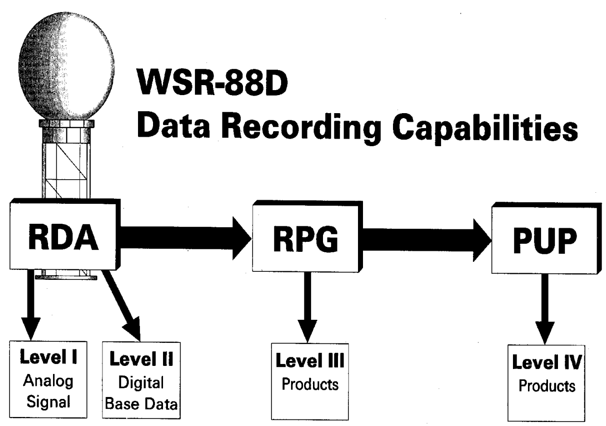


Figure 27 - WSR-88D Data Recording Capabilities (Crum & Alberty, 1993a)

While the hardware and programming here is much more complicated, the discussion in the following sections will reveal that a) Level 1 data usually remains at the radar location unless it is recorded and sent away for analysis, and b) Level 2 data is what can be considered the “raw” data. It is composed of six “moments” that capture discrete measures, for example, the strength of the energy reflected back (reflectivity) or a measure of whether or not a phenomena is moving toward or away from the radar as the radar scans the sky. It is this data to which algorithms are applied via the RPG to process and interpret these data, resulting in c) Level 3 products mostly in the form of images, although some consist of data, and d) Level 4, more processed depictions of weather phenomena. The rough distinctions between Level 1, 2, and 3 data will be helpful in the discussions ahead.

Making Value: Instrument, Algorithm, Archive, Access

Organization of this Analysis

One approach to considering influence on data and related policy, as well as the impacts of mutual dependence on data sharing between government and external users, is to use the lens of the "information lifecycle." That is, to consider the life of the data itself from "cradle-to-grave," looking for influence along the way. The information lifecycle is not a novel concept. Federal Office of Management and Budget Circular OMB A-130 (OMB. 2016) frames government information policy using this lifecycle. And, NOAA has published the Environmental Data Management (EDM) framework (EDMC, 2013), a more detailed version of the information lifecycle for environmental data.

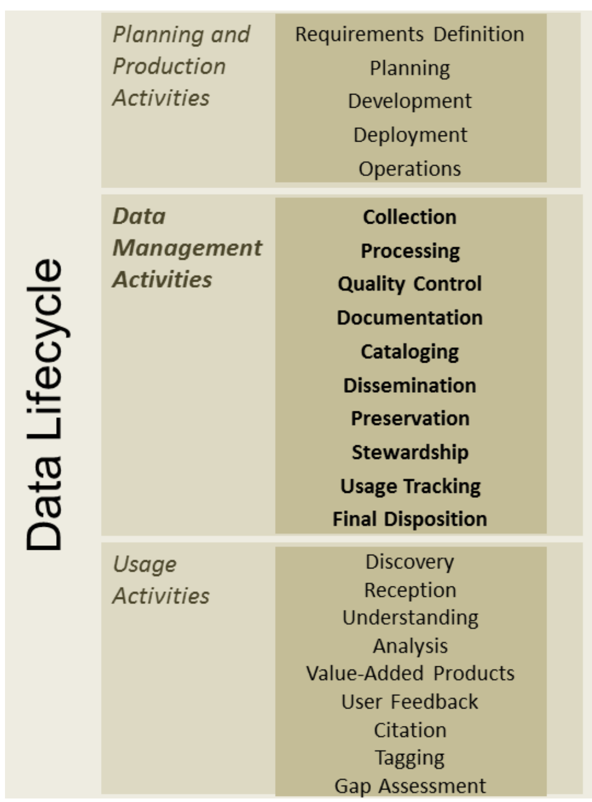


Figure 28 - EDMC Data Lifecycle (EDMC, 2013)

In OMB A-130, “Information life cycle” is defined as “the stages through which information passes, typically characterized as *creation or collection*, *processing*, *dissemination*, *use*, *storage*, and *disposition*, to include destruction and deletion” (OMB, 2016). NOAA’s Environmental Data Management Framework 1.0 (EDMC, 2013) – see Figure 28 above - is a more detailed elaboration of this lifecycle and recognizes the idea of Data Management activities as part of it. This model groups the life cycle into three functions: *Planning and Production Activities*: *Data Management Activities*; and *Usage Activities*, each made up of related stages. While this is also helpful as a conceptual frame for analysis, Usage Activities sit outside our present focus. And, for our purposes, requirements definition, planning, development, and operations could actually be considered a “sub” life cycle for each Data Management activity shown. To further simplify both these frameworks for our analysis of NEXRAD, I have abstracted *collection, processing, preservation,* and *dissemination* (or rather “distribution”) from the EDM framework, paralleling the stages of creation / collection, processing, storage (or “retention”), and dissemination in OMB A-130. In applying this to what we have learned at a high level about NEXRAD weather radar data, I arrive at four lenses for analysis: **Instrument** (used for creation / collection), **Algorithm** (processing), **Archive** (preservation / storage), and **Access** (dissemination / distribution), with the idea that there are planning and data management activities of the types in the EDM framework throughout. From the discussion of information construction in the introduction to this research study, I assert that there are meaningful and important choices in each of these stages that can impact the content, structure, and availability of the resulting data. In turn, to the degree there are choices available, each stage presents an opportunity for influence. Given that NEXRAD was designed and implemented over thirty years ago, access to details about and individuals involved in that process are, by definition, limited.

However, using this approach will also allow me to illustrate the kinds of possibilities available in each stage with regard to shaping the data, even when evidence is hard to come by, as well as clearly present those elements where such evidence is readily available. This should also be helpful in thinking about the role of path dependence in decisions about technology in other areas as the analysis continues, and their consequent role in shaping the arc of benefits, influence, and mutual dependence on the secondary use of this information.

Origin and Oversight of the NEXRAD Program

As context for the discussion of the history of the NEXRAD program and stages in the lifecycle of the data it produces, I begin with a brief overview of the organizational context in which the NEXRAD system was developed and implemented in the 1980's and 90's. The project was undertaken by - and to this day is the joint responsibility of - three federal agencies: The Department of Defense (each branch of the military has interests here, but especially the Air Force and Navy), the Department of Commerce (NOAA and its component, the National Weather Service) and the Department of Transportation (where the Federal Aviation Administration resides) (Whiton, et. al, 1998). The implementation of NEXRAD was overseen by these three government agencies through the NEXRAD Joint System Program Office (Crum & Alberty, 1993). We will see in later discussions that this Joint System Program Office eventually hands over these responsibilities to a successor organization, but one in which all three organizations remain involved.

Instrument (Collection)

The limitation on access to information and interviews about the history of NEXRAD is most evident in the case of the design and development of the initial system. NEXRAD was a large and visible project costing over $1B (GAO, 1991) that involved significant resources and planning. As one might imagine, determining the placement of over 120 radars across the nation was a big undertaking, involving engineering studies and site visits. For example, a map was created from nighttime satellite images of the United State to analyze lighting patterns to better understand population distribution in areas that the radars would provide early warning and devise optimal coverage (Leone, et. al, 1989). It seems there would have been significant challenge for a commercial user of the eventual data from the system to have much influence on these locations. However, the point here is, again, that these decisions ultimately influence the nature of the data that will be produced and they are products of their time and the technologies available to make these determinations.

Development of the NEXRAD system also required a long period of design, development, and testing of technologies, modules, and display prototypes, as well as, evaluating "operational effectiveness and suitability of pre-production radars developed by the competing contractors" that involved the National Severe Storms Laboratory (NSSL) – a part of NOAA that we shall encounter again later in this chapter, and other government organizations. Upon testing by a "160-member tri-departmental team" (Weyman & Clancy, 1989) the contract to build the NEXRAD system was awarded to Unisys over the only other bidder, Raytheon – a company who had been the developer for previous weather radar systems (PRI-31). A brief review of the 231-page November 1, 1991 NEXRAD Technical Requirements document (NEXRAD JSPO, 1991) shows it to be very specific in its requirements for the data and products to be produced. These decisions about requirements may have been open to influence of various types and purposes, given the size of the project and the number of people and organizations involved, but I do not have evidence either way.

While it is difficult to uncover the internal machinations of what happened so long in the past, the impact of negotiations among different parties at this time, technologists vying to implement their vision for the best way to address the challenges of designing new functionality, and businesses seeking to win the competition for the development of the system, government's understanding and their requirements across three agencies for what would be needed for accuracy, reliability and flexibility in the future – all appeared to have played a part. Similar processes seem possible in design of computer systems that collect or produce other data in our study so far – disparate design of restaurant inspection systems, systems that maintain driver records, or even the genealogical records that were produced long ago using one technology (paper) that set the stage for the value of another set of technology services (digitization) that can traded by industry to government to overcome what are now considered deficiencies in the data format produced in the past. It should also be noted that while NEXRAD was developed many years ago, the choices made in the development of these systems, often purchased as commercial-off-the-shelf systems by government for these purposes, may be just as opaque.

While identifying direct and specific influences of business and government in the design of such a major project so long ago is, from what I've seen, a task that would require its own dedicated research study, highlights from a conversation with one of the programmers on a key piece of the NEXRAD system – the "programmable signal processor" that produced the "moments" of raw data from the radar scans to which algorithms are applied that forms the base for NEXRAD products used even today - provide a flavor of the many factors that went into setting the capabilities and limitations of the system. The Level 2 data that these moments make up (three more have been added for dual-polarization, but these are still there) and direct access to them by government and the private sector will play a large part in later discussions in the

Access stage of this lifecycle analysis. The following sections introduce just some of the influences involved.

Government Preferences and Requirements. In the lead-in to this section, I noted there were three government agencies involved in setting the course for and overseeing the project. Here is the developer's on-the-ground take:

> The program actually had three customers from a government standpoint with slightly different needs. The National Weather Service being for general storm warning for the public as well as over all other weather monitoring and NOAA and their mission. The FAA, (part of) the Department of Transportation, was especially interested in storm warnings as affected airport traffic. And in particular, the en route part of airline traffic. Other radar systems that keep closely in track what is happening right around an airport. But it terms of having warnings across the country to know where planes might be flying through and warding them off of storm areas, that's a certainly part of the original mission. The Air Force's primary interest was to know when they had to scramble airplanes off of a base because airplanes exposed on the ground, for example, in a hail storm, or something like that are pretty vulnerable. It's very expensive to put hundreds of airplanes in the air just to get them out of the way. That was actually the Air Force's, as I was told anyway, primary interest in the program. (PRI-31)

One can imagine then, that establishing the requirements for this system, even on the government side, involved some tradeoffs and negotiation. This is supported by a journal article written about those times (Whiton et. al, 1989) that stated: "Differing agency requirements and program management methods, fiscal pressures, and some distrust among the agencies made the first year difficult."

Industry Competition. As noted earlier, the initial phase of the development project was set up as a competition between Sperry (Univac) and Raytheon to come up with their best design for the new capabilities. This developer worked on the Sperry side – the eventual winner - although he noted that the company went through name changes and consolidations during the project. Raytheon, however, had developed and provided the government with previous generations of the technology. In commenting on the situation, the developer observed:

> ... In some sense they (Raytheon) had been the incumbent and politically, they certainly had some friends inside the National Weather Service who would, presumably, have good experiences working with them over the years. And didn't necessarily want them to be replaced. And we were sometimes aware of that. There was a big contract and there was fair amount of politics involved. At that level, didn't affect me at the engineering level but these things were going on. (PRI-31)

Technological Requirements, Innovation, Limitations, and Time Constraints.

Recalling the diagram of the NEXRAD system in Figure 27 above, the “programmable signal processor” this individual worked on was associated with the first major component, the Radar Data Acquisition (RDA) unit:

> Actually there was a first set of electronics as part of the radar that basically converted the analog to digital and did some preliminary filtering on it. At that point our piece of the signal processor that we were working on, it did a few things, but its primary function was in creating what are called the “moments.”
>
> Then (the system would take) this basic, moment information and that's where the algorithms would actually detect storm patterns, etc. What would get built. So you could think of the things coming out of the RDA as the things... The pictures on the screen? We would generate the first levels of those pictures. The ones that would look like a circle and that kind of a map.
>
> I believe that the moment algorithms were left up to the contractors. That was part of the contractors' responsibility. This got into some disputes later on. In principle, any meteorological algorithms that actually had to do with how weather was constructed and such were actually the government's responsibility. Some of that stuff would have been as they would say above my pay grade, but what I would say is that the fundamental capabilities that the radar had to have in terms of sensitivities and accuracies and things like that, they were set by the government scientists. That came down to us through the Joint Program Office, but those are part of the basic requirements.
>
> And those are very important. We had to, whatever algorithms we did for our moments, it all had to provide... A lot of it had to do with what was going to be physically in the radar. It ultimately drove things like how big the dish had to be, which was a big expense factor. It had to be 28 feet instead of 24 feet for example. Some of those, you get down to the science of what you need to do to meet those requirements. The meteorologists, the radar scientists, I would think like the National Severe Storms Lab (NSSL) were putting out the basic requirements to get the capabilities they needed. It was up to the contractors to figure out how to make that happen up through what the RDA would produce.
>
> (PRI-31)

In a later section on Access in the information lifecycle, we will learn more about the ongoing role of the NSSL and their relationship to the private sector. Here, though, while it might seem counterintuitive, we see that the development here started with determining the data that was expected to be received, then ensuring it could be processed appropriately – and then designing the dish to collect it:

> We had a simulated data source that we were able to pump in to our signal processor and get the outputs...and then over to an RPG where they could actually then see a simulated map. As a matter of fact, all it was showing was rings of constant data because we couldn't build a simulator with enough memory to show a whole sky full of actual weather radar data. We were basically demonstrating that the competency chain actually worked and that the algorithms were working. (PRI-31)

These stories demonstrate not only the critical interaction between science and implementation in this design, but factors such as cost constraints that came into play. Of course, so long ago in the history of computing (and keeping in mind that today will be “so long ago” someday), there were other limitations. He continued:

> When we were doing this thing in the '80s, our high performance computing thing that we built off our programmable signal processors, it was really state of the art, but nowadays the things we needed a whole rack of equipment to do they can do on a couple boards. (PRI-31)

And, they were, like most projects yet today, operating under significant time constraints:

> We had 10 and one half months from contract turn on until we were supposed to have actually the whole programmable signal processing running with its hardware, with its software, with all of its applications, moment generating stuff.
>
> In those days in particular, there's no lead time for parts. We were starting from ground zero…They designed the processor basically from scratch. Some of the concepts were based off some things that had happened before. But these were days before microprocessors were just becoming to come on line, but they were not capable of doing the kind of processing we needed at that point.
>
> The computers were actually being built from scratch from individual electronic components. It required designing and building eight new board types. And in those days, it took months to turn around a printed circuit board. Things that nowadays get done basically in days. It took months. We basically ended up having about six weeks from when the hardware was marginally workable - which is to say the first copy rolled in and

> still had lots of bugs in it - to when we actually had all the software up and running on it. (PRI-31)

As we come to the end of this section, we close with an anecdote from the developer that demonstrates how these forces of competition, innovation, government requirements vs. their interest in programmable vs. "hard-wired" features that would allow future flexibility, and even path dependence in the orientation of management and resources in the companies involved came together in the key part of the system he worked on, the "programmable signal processor":

> There was interest from the government side in having the capability to adjust things like the moment calculations even though the contractor was not on the hook to do it…So it's not a requirement for the government. But it was considered a plus that things were going to be more programmable. Now in terms of this – it gets into some corporate politics.
>
> If you've never worked inside big corporations, you might know that sometimes fights inside a corporation are bigger than fights between corporations. Eagan (the branch in Minnesota) had been that part of Sperry Univac, which had originally been UNIVAC and in this particular case it was a defense parts of UNIVAC, was primary a supplier of computers for the Navy and the Air Force. Specialized computers for military applications. What we did with programmable signal processors was a little bit outside of the mainstream for them. There had been some technology derived from the consideration of a Doppler system for the FAA that had not happened. That would have not been a weather radar as such but it drove some processing development.
>
> To put it mildly, our corporate president at Sperry Gyro in Long Island did not want the signal processing part to be programmable. They had been traditional radar developers and they wanted to basically hard-wire, not really what you would call hard-wired into code, but to have the electronics just be specialized electronics. It didn't have anything that you would think of as programmable, it would just be a fixed electronic outputs that would do the digital processing in a completely tailored way so the algorithms would, in effect, be completely built into the hardware.
>
> That was really their preferred technical approach, *but they knew some people in the government wanted to see something more programmable*. They reached out to us to put up a straw man. The real intent was to knock it down and not use it. As a matter of fact, we came up with new concepts in the first year that attracted government approval. *They found they were getting plus points when their evaluators came around because they were pursuing this.* [emphasis added]. And eventually, as we went in to the next phase of the contract, they made us an official part of the system.

(PRI-31)

While these anecdotes from the early days of NEXRAD development help us gain insight into technology's role in dependence and influence, in considering private motivations associated with this phase of data 'construction" in current projects, the comments of a high-level executive at a major U.S. weather company about standards, raw data, and the key role played by the design of the instrument are also instructive:

> Standards are almost always a value reducing exercise. I should be clear about this. Taking data from its native form is dependent upon the precise technology that was used to create the data in the first place. The optical sensor on the satellite or the structure of the data inside the numerical weather prediction model software code set or whatever, that's the native form of the data in its purest and most valuable - it contains the most value. Anytime you transform data, you're almost always reducing the net value of the data. You're reducing it in the sense that it's lost information.
>
> You may actually, in doing that, make it more easily accessible to a broader set of applications. The net value you might get from that is increased, but for those who want the epitome of the value from the native information, standards is not something you want to do. The space I personally play in is I want that raw native data because I can figure out ways to extract value out of them. (PRI-11)

From these statements we can take that industry is well aware of the potential commercial value produced by choices in hardware and software design (and standards), an interest that is further reinforced in industry's attempt to gain greater participation in NOAA's design processes that was included in the SAB recommendations in Chapter 3.

These observations make obvious the role of design and placement of instruments in the construction of the resulting data, and that choices here – about standards, technologies, or even software routines - are places where the exercise of influence could set the course for possible uses further downstream, including by the private sector. The next step in the lifecycle – Processing – shows possibility. Industry uses these products to customize and add value to them so they may have an incentive to influence the nature or priority of the algorithms deployed to produce additional commercial value. But regardless, development and implementation of these technologies over time further affect what can – and can't (for now) – be done with the resulting data.

Algorithm (Processing)

In the Beginning: Influence and the Algorithms as Originally Delivered in NEXRAD. NEXRAD was implemented with a defined set of algorithms and products. As briefly discussed earlier, these algorithms are computer programs that have been designed to process "returns" of signals sent out by the radar, collected in the form of Level 1 analog data, by applying mathematical operations to remove, enhance, combine, or otherwise derive value from them. With regard to direct commercial influence in the design of the original set of algorithms implemented in the system, this excerpt from the NEXRAD Joint System Program Office Next Generation Weather Radar Algorithm Report (NEXRAD JSPO, 1985) provides the official government perspective:

> 1.1 Algorithm Philosophy
>
> The NEXRAD system acquisition process follows the guidance given in Office of Management and Budget Circular A-109. The A-109 approach is designed to make the most of the technical expertise that exists within industry. However, *the Government has determined that the meteorological and hydrological analysis techniques necessary were only available within the Government. Thus, the Government has assumed responsibility for providing those meteorological and hydrological algorithms to the system contractor* [emphasis added]. The contractor will implement these algorithms into a real-time system. (NEXRAD JSPO, 1985).

As context, OMB Circular A-109 (since rescinded), in turn, essentially directed agencies not to make their requirements overly prescriptive, as expressed in the following provision:

> a. Express needs and program objectives in mission terms and not equipment terms to encourage innovation and competition in creating, exploring, and developing alternative system design concepts. (OMB, 1976)

… a provision that might have created more opportunity for influence by industry. However, although this statement seems to direct agencies to give contractors flexibility in developing equipment, the bottom line is that it would appear to have been effectively overridden by what followed: A 900+ page specification for the initial NEXRAD algorithms (NEXRAD JSPO., 1985). The clear and prescriptive nature of the specifications leads one to believe that industry influence on the algorithms was limited. This is also generally supported by the observations of the programmer interviewed in the last section who saw the requirements primarily determined by government scientists.

Another reason to believe that industry had less influence over the development specifications for NEXRAD algorithms is that the project was subject to independent audits during its implementation. A U.S. Department of Commerce Office of Inspector General Report from February 1993, in addition to noting ongoing contractual problems with the vendor Unisys that resulted in congressional hearings and temporary stoppage over the quality of the work, assesses whether the software (then in excess of 400,000 lines of code) could be successfully maintained by the government on an ongoing basis (DOC, 1993). While their overall conclusion was that this could be achieved, they issued a number of recommendations, noting that “More visibility is needed into NEXRAD algorithm changes” and consequently recommending a more formalized process for this to occur going forward (DOC, 1993). They explicitly stated “[W]e are concerned that decisions regarding algorithm changes after the OSF assumes its support responsibilities will not be sufficiently controlled” requesting that the OSF develop an annual plan for “modifying and adding algorithms” that would “present the *rationale* [emphasis added], technical approach, and cost and schedule estimates for the proposed changes and additions . (DOC, 1993).

These concerns were an outcome of an earlier (1991) inspection that accompanied a renegotiation of the contract with Unisys (delivered through its subsidiary, Paramax) that found

"[P]oor software and engineering practices" (DOC, 1991). From a review of both these documents, the concerns appeared to be less that there was an attempt to modify the algorithms to perform additional or different functions, and more that there was "little engineering of quality during the design phase" and that "software developed in this fashion has a high risk of being error prone and difficult to maintain" (DOC, 1991). A NEXRAD software assessment team drawn from the Air Force Electronic System Division was formed to review Unisys software development in February 1991 and issued recommendations for improvement.

For our purposes, these activities speak to the high scrutiny this work received with regard to its quality by the government in terms of, assumedly, its responsiveness to the JSPO algorithm development and other software requirements, limiting discretion available for new or different functionality to be included without detection by the government. As we shall see in the next section, the rigor of the resulting processes put in place at the Radar Operations Center and the absence of parties external to the tri-agency sponsors in its governance (industry has no representation in any of the resulting processes) also serve to make direct influence more difficult going forward.

Mission Focus: Governing Algorithm Development, Prioritization, and Implementation. While a part of NOAA organizationally, the NEXRAD Radar Operations Center (ROC) supports the NEXRAD program and is staffed, funded, and governed by all three sponsoring agencies. Along with planning and performing maintenance and upgrades for the radar-related hardware and underlying operating software that make up the NEXRAD systems, it is here that the algorithms that make the data and products coming out of the system are designed, tested, and implemented in the Radar Product Generator (RPG). As just described, both the experiences of the NEXRAD implementation project and outside audits reinforced the implementation of rigorous oversight and processes to manage decision making for all support operations – and the processes related to algorithm development are no exception.

The identification, development, and implementation of new algorithms (and changes to the existing system) occurs in a tightly governed software development and configuration management process that requires multiple levels of justification, review, and approval, including technical review and priority / staging of their implementation. While the following discussion, by necessity, is not comprehensive, I intend it to provide a general view of the processes involved by way of illustrating the challenges to direct influence by commercial users on the resulting products/data. A subsequent section discusses incentives and rationale (or lack of it) for such influence on NEXRAD algorithms outside this process.

In brief, changes to the NEXRAD software baseline (the release level reflecting all the versions of the software components making up the production system) may have different origins. Some come from Radar Operations Center engineers, some from research within one of the three agencies that sponsor NEXRAD, or from the National Severe Storms Laboratory, a part of NOAA – but none come directly from external parties. The situation is described by an official at the Radar Operations Center:

> The agencies that fund us, they have a portion of their funding that is geared towards improvement, product improvement. We manage that money by partnering with the National Severe Storms laboratory. We partner with NCAR (National Center for Atmospheric Research – an academic consortium), and we partner with the FAA, which the FAA has it as a link to Lincoln Lab (federally funded lab at MIT). These are all experts in research and development of radar. We are not funded to do research; we're funded to maintain the network and run it ... and prevent it from failing. We have to create a partnership through a memorandum of agreement with research entities. They take that money, and they do the research.
>
> Now they vet that research at meetings like the AMS (American Meteorological Society). That's the peer review side. National Weather Association (an industry association), all of the ... the AMS radar conference. They go on and get peer review, which is open to the

> private sector. That's where there's an interaction at that level in the private sector with our research arm. (PRI-32)

So, there appears to be little direct solicitation or application of commercial influence at the Radar Operations Center. However, routes of possible influence by external groups on algorithm design, such as those mentioned in the comments above, will be addressed later in this section. To start, however, we should understand the types of development that occur. Changes to the NEXRAD software are categorized as *enhancements, bug fixes, maintenance, and obsolescence* (ROC, 2009). Enhancements are further divided into major and minor enhancements, with the latter having no external impact, funding, or hardware changes. The algorithms we are concerned with here that essentially create new data fall into the category of major enhancements. Major enhancements are expressed in the form of a Configuration Change Request (CCR) and require the approval of a Technical Advisory Committee (TAC) and a Software Recommendation and Evaluation Committee (SREC). One role of this committee – again, made up solely of government managers – is to determine the readiness of an enhancement to be included in a future release of the software and its impact. As part of this evaluation, two questions required on the “WSR-88D Algorithm Process Template” demonstrate the close connection between these enhancements and the program requirements of the tri-agency sponsors:

> 2. Operational or System Requirement to Be Met with Change. Include either new or existing agency requirements/goals/strategic plans addressed by this proposed change. Identify which NEXRAD agencies and types of users will benefit. (ROC, 2007)
>
> 9. IV&V of Scientific Validation. Describe what validation of the technical goodness of this change has been completed and who has reviewed the validation work (e.g., NEXRAD Technical Advisory Committee, published scientific journal). (ROC, 2007)

So, as noted, different governance groups become involved at different stages of the process, some via triggers related to impact (as with the major enhancements above) or cost. There is a Change Control Board that manages the flow of changes and the entire structure is overseen by the joint NEXRAD Program Management Committee, made up of representatives from the National Weather Service, Air Force, and FAA that provides oversight for "budget, policy, resource commitment, (and) management guidance" for the NEXRAD program as a whole. The following slide from a NEXRAD software engineering presentation demonstrates the administrative overhead of initiating an algorithm change:

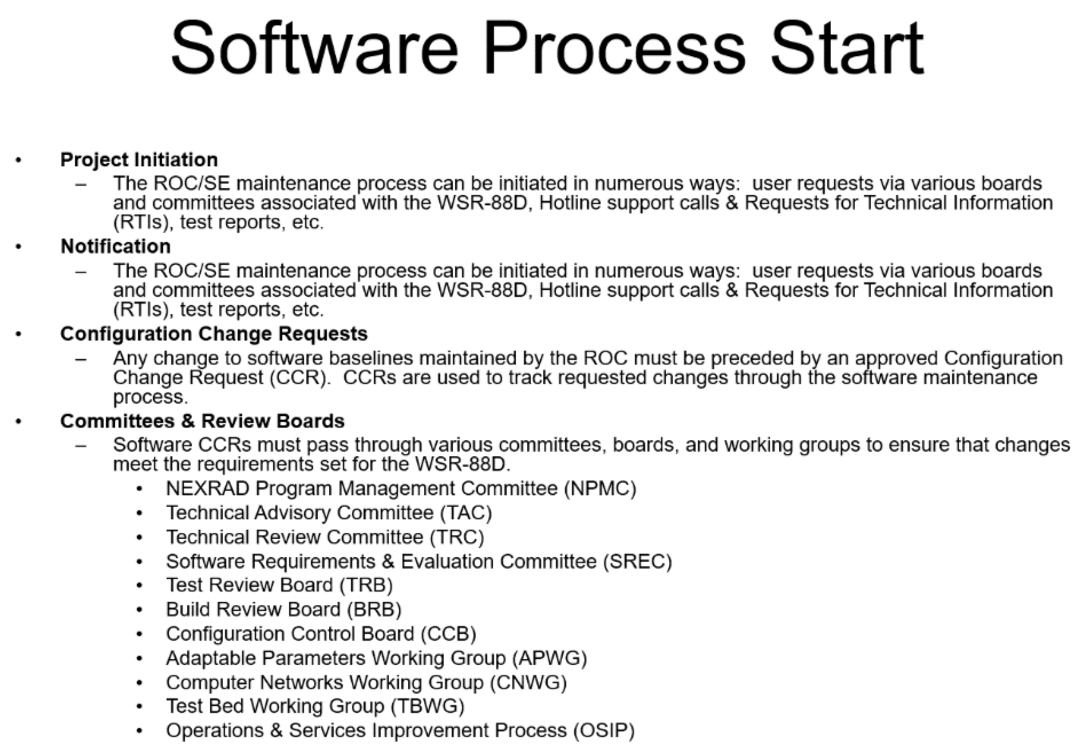


Figure 29 - Software Process Start (Approvals) (ROC, 2009a)

While the executive above noted that development of these algorithms may be discussed in various meetings of industry professionals – it is not a secret process – it is clear, as he states “I take my requirements from the agencies that own the radar” (PRI-32). All said, it seems that the possibility of direct influence here is limited.

Macro-influence through Governance: NOAA/NEXRAD Planning and Priorities.

In the introduction to this section, I mentioned that the NEXRAD system was determined by the Office of Science and Technology Policy to be one of the most valuable "observing systems" to a wide range of parties. The executive at the ROC interviewed above referenced a case where an individual lobbying Congress resulted in the construction of a new NEXRAD location in the State of Washington: "We don't generally see them introduce themselves into a technical solution (but)…That's an example of how a public or private sector person can go to Congress and levy a requirement on us" (PRI-32). In addition:

> There's a couple different scientific-based expert panels that work with Congress because Congress is the arm that directs us. They'll be commissioned by Congress to look at the National Weather Service and critique the performance of the agency. They can directly reach in and influence the course of, for example, the NEXRAD program. The National Weather Act reached in and looked at radar. So they're not all government people, they're a variety of academic people, et cetera. But they will work for Congress, and then Congress will provide direction, and it flows down to us. The National Research Council… So they're not all government people, they're a variety of academic people, et cetera. (PRI-32)

The influence on shaping NOAA policy through reports from external groups like the National Research Council and the National Academy of Public Administration, or the Science Advisory Board played a large role in the discussion of industry influence in Chapter 3. It should be noted, then, that there have been reports specifically focused on NEXRAD over time. The National Resource Council's (1999) *Enhancing Access to NEXRAD Data - A Critical National Resource* contained recommendations, among others, to improve the archiving and availability of NEXRAD Level 2 data for research purposes (which, as we shall see, was implemented and created wider availability for industry) (NRC, 1999). A previous report, *Assessment of NEXRAD Coverage and Associated Weather Services*, evaluated the implementation of the radar (NRC, 1995), and another, *Weather Radar Technology Beyond NEXRAD*, laid out recommendations for its future (NRC, 2002). These, coupled with numerous other reports on aspects of the weather service, its programs, and reorganization, confirm significant interest in the topic and reconfirm the idea in an earlier chapter that these external advisory groups are, by their nature, designed to influence programs in the form of recommendations and may reflect some industry interests. For example, 2003's "Fair Weather" (NRC, 2003) that explicitly acknowledged interest in NWS decisions, as seen in the recommendations the report outlined in Chapter 3.

In this section on influence on algorithm development, rather than exhaustively tracing recommendations that might touch this part of the lifecycle (see the Archive and Access sections for more on this subject), I will leave it that these groups are set up to provide legitimate influence and advice that, in some cases, explicitly focuses on NEXRAD.

There are, however, two aspects of influence that are internal to NOAA and somewhat closer to its day-to-day management, that are nevertheless external to Radar Operations Center day-to-day operations. The first is the NEXRAD Strategic Plan (Clark, 2016). While it was created to reflect the priorities of the tri-agency governance structure discussed earlier – and does not reference NRC report priorities or recommendations - it includes an acknowledgement of "commercial and business enterprises" as a stakeholder in "program decisions, activities, or outcomes" (Clark, 2016). In turn, we see something similar in a document outlining NEXRAD functional requirements submitted for approval to a group discussed in Chapter 3, the NOAA Observing Systems Council (NOSC). The NOSC, among its other responsibilities, assesses NEXRAD's value among other observing systems. In the document approved by the NOSC (ROC, 2015) that contained NEXRAD's functional requirements through 2030, the scope statement indicates it considers "multiple areas" …including "support to other Government agency and private sector partners." (ROC, 2105). While the bulk of the requirements are for the Radar Data Acquisition and processing component of the NEXRAD system, in a section devoted to "additional discussions to help place these requirements within the context of overall radar information use by NOAA/NWS and other users" (ROC, 2015), they list many users of the system that include "Private meteorological sector (e.g., commercial weather companies, television stations, energy generators, drought monitoring, fire risk, water resource management, etc.) (for use in) tailored forecasts for specific customers." With regard to this list, they go on to state:

> These users have developed an extensive infrastructure to acquire and process the NOAA/NWS radar data. It is critical that future plans for additions to, or modifications of, NOAA/NWS radar functionality consider these users' investments, minimize disruptions to their operations, and minimize the need for costly modifications of infrastructure. (ROC, 2015)

Overall, then, external stakeholders are acknowledged and, in the preceding example, considered as part of NOAA decision-making in the vein suggested in the interviews referenced above, although I have seen no explicit evidence of this occurring in my research. From Chapter 3, however, in support of the idea that mutual dependence exists, we did learn from a conversation with a member of staff in NOAA's Technology Planning and Integration for Observations section that NOAA used its public value determined by the NOSC and OSTP to lobby for additional funding within the organization (PRI-33).

As for peer influence exercised through the meetings of the American Meteorological Society (AMS) , this is difficult to quantify. There is a section of this group with its own "sub-conference" held in association with AMS annual meetings that pre-dates NEXRAD (it began meeting in 1985). Called "Interactive Information Processing Systems (IIPS)," it has been renamed but is still held today. A review of conference agendas – the meeting also covers hydrology and oceanography – does show frequent presentations on algorithms under development (see agendas: AMS, 1999; AMS, 2001; AMS, 2004; AMS, 2010), including, for example, in 2006 where Weather Services International (later The Weather Company) presented on algorithms they had developed for turbulence and related to airport conditions (Sousounis, 2006). Private sector influence on algorithms included in NEXRAD would be subject to some of the procedural barriers already discussed at the Radar Operations Center. But, as we are about to see, there is another, perhaps stronger, disincentive – competition.

Commercial Incentives and Barriers in Algorithm Development. From the preceding review of the history and current practices in algorithm development at the Operational Support Facility / Radar Operations Center, it appears the complexity of the processing required to create NEXRAD products and the software engineering and governance processes required to manage its design and development left no avenues for direct influence on the content or structure of the resulting data. However, this in itself is a finding, in that certain types of data may lend themselves to more or less rigor around the processing phase of the lifecycle, and, in turn, it seems possible the opportunity for influence might vary accordingly. Other aspects of this situation are also of interest, but to engage them, a fact that will emerge in a later section on Access must be brought to the fore here. The algorithms under discussion above are applied not to Level 1 data – the analog data produced by the radar that is not retained – but instead to Level 2 Digital Base Data, the "moments" that are processed to create products (Level 3 data) for business and the public.

The key here is a fact omitted so far, but that will be addressed in more detail in the section of this chapter devoted to the Access phase of the information lifecycle. Both Level 2 and Level 3 data are available real-time (and have been for more almost two decades) to the private sector, free of charge. Armed with this knowledge, we are then led to another salient aspect of the situation with influence on NEXRAD algorithms – <u>the private sector has access to the same Level 2 data as the government uses to create Level 3 products to which it can apply its own</u>

proprietary algorithms to create additional or improved products for sale.

When this factor is considered, the rigor of the selection and development process for algorithms as a barrier to influence may suddenly seem superfluous. As the Radar Operations Center official put it:

> The predicate here that industry tries to tell the weather service how the radar should produce information is probably not viable because Industry wants to take the data we have and then make it something that they can make money from. If they have an idea, they generally bring it in through the research area, which is an AMS conference, a radar conference. Industry comes into those things all the time and shows us what they have. That's their forum to influence what one of our agencies might see as a requirement they would fund. (PRI-32)

So, far from being incentivized to enhance the processing via algorithms of the data the government is producing – data that is free to all competitors in the weather services market – companies are instead likely to develop and apply their own algorithms to the Level 2 data to create proprietary products, giving them a competitive advantage in the market. This idea was expanded upon by a former Radar Operations Center software engineer who developed algorithms there and then moved to doing similar work in private industry:

> I think the thing that would surprise me about that (industry influence on the design of algorithms at the Radar Operations Center with an eye toward influencing the data produced) is that with the level of the way this thing works, all these private sector companies look at themselves almost as VARS - Value Added Resellers. *So, for them to come in and try to lobby for the government to actually develop those algorithms, I just don't know how much that would be in the private sector's best interest. Because then now everyone's on the same playing field* [emphasis added]. If anything, what I had heard from the ROC, and this was just hearsay - this was never anything that came down from the top - *is that there were already a lot of private sector companies that didn't like how much development we did* [emphasis added]. (PRI-34)

And now, we have come full circle, back to the policy negotiations in Chapter 3 and the concept of "Boundaries of Convenience" (as determined by industry) that prevent NOAA from straying too far in developing its services in ways that might compete with the private sector. As we shall see later in the section on Access, there was a time when Industry did not have access to the Level 2 data – it was not widely distributed even within the National Weather Service. But, once it became available, industry moves the public-private line closer to NOAA to limit their activities, characterizing service enhancements as competition where it suits them. The algorithm developer continued:

> It did kind of cause some rifts sometimes. The meteorology community is very small so you always hear rumblings and grumblings. "Oh, I can't believe that they're (NOAA) going to do QPE (Quantitative Precipitation Estimation). They should just leave that to the private sector." Or "They're going to do hail size detection. They should leave that to the private sector." You know unfortunately at the ROC we struggled to really maintain very apolitical stance. Our goal was purely to support the weather forecast offices. And, you know, when they come and say, "In order to save lives and property" which is their mission, right, they need these algorithms. And, so "Private sector kind of be damned if we're not going to.".. That's our mission is to support them.
>
> I think a lot of it is the motivations (between NWS and Industry) are very different. The motivation of the ROC to support it is because of operational meteorology and so usually it's a WFO (Weather Field Office) or somebody like that who may dream up an idea and say, "Hey, how do we get this in there," or, "How can we do this and get this a little better?" The problem is it's super slow for the ROC to respond to that because it's got to go through all that. There's a lot of tap dancing going on there. The private sector is much quicker to respond obviously to these things, but with that usually comes specialization of the algorithms. The algorithms that these large companies develop are vastly different and typically very tailored. A lot of these companies are maybe like agriculture companies, or transportation companies, so they're looking to do very specific processing. The motivations are very different and often times you (for example) get six companies working on a quantitative precipitation estimation algorithm to know how much rain fell from just the radar.

(PRI-34)

Here we have an interesting impact of access on influence. In a world where government data is open to everyone by policy, a commercial user is in competition with other companies who have the same resource. In this case, if NOAA adds a new algorithm to create products that serve substantially the same purpose as those that I develop with my own, suddenly there is little reason to pay for my product. So, one must keep in mind that the private sector playing field is not level, and the winners like it that way. The algorithm developer has another observation as well. He stated that the Level 2 data that's distributed has not continued through the quality control processes that occur later in the NEXRAD system, so it requires a significantly higher level of skill to work with, resulting in the case that:

> There are a lot of companies that really leverage only Level-3 products just because maybe through business's decisions or knowhow of their individuals, it might not be worth it for them to develop the algorithms. (PRI-34)

So, while it remains possible that some companies would try to influence the process, competition among companies would work against this, given that all would have equal access to the resulting products (and potentially, the benefits) via the federal government's open data policy if they were successful in getting NOAA to adopt the change.

Influence and Data Structure. The structure of the various levels of NEXRAD data is essentially determined by a combination of hardware and software. Algorithms, as noted previously, are applied to create Level 3 products. And, our research questions include an interest in how (mutual) influence might affect the structure of data**,** that is, the formats in which is it delivered. Formats / standards are important for a reason stated by a weather executive earlier in this chapter – they to some degree reduce value in that they may shape the meaning of data through application of categories, or remove some details that may be viewed as extraneous or a barrier to data sharing, potentially eliminating an attribute that could be critical for another user.

In the course of talking about NEXRAD, it comes up that while standards are important to data sharing, systems are / were built naturally anticipating who would use the data and how. It is hard to imagine that engineers could, for example, anticipate all the eventual uses and parties involved in products created by NEXRAD over the years, and thus seemingly impossible that they would have anticipated their needs from a technology and standards perspective. An engineer who has worked nearly 20 years with NEXRAD data, including in government research, with private sector companies on weather-related science using the data, and now at one of the Big Data Project collaborators, recounted:

> [I]t used to be that radar data was collected at the radar and then it was essentially used only by the weather forecast office associated with the radar. Then it changed with the advent of Internet 2 in the late nineties where you could suddenly put the data on the Internet and have it distributed. Basically, it was meant for use only by probably one system. And now, we are essentially changing it to be used by a whole bunch of people.
>
> So then, we (NOAA) built the systems to collect the data from across the country, 140+ radars, and then bring them into a central location and then be able to basically process them in real time. That again, with the computing infrastructure of those days, that was pretty serious amount of processing that we had to do. Of course, the data, in many ways, are from that era. The whole data formats and everything were designed in the era of, you're basically sending the data over T1 lines. Like, really no bandwidth. How do you get the data over to the weather forecast office? No one really cares about how you get the data beyond that, right? It's just very limited distribution of the data. (PRI-35)

Here we see a concept that may apply more broadly. There are certainly many legacy (older, perhaps not using current technology or standards) systems in place in government today, with standards for data and technology using an internally-focused design based on a well-defined, limited use. He continued, explaining the dilemma they faced as the standards of external users changed:

> And, of course, we have data going back to 1995 now in the NCDC (now NCEI) archives that is in an old format. So, when Polarimetric radar (Dual-pol) came around, which was in 2012 or 2013, the question was "Okay, what format do we put it in?" and there was a lot of push at that time to make this a format that is self-descriptive, modern, that anybody can use. But, then there is all of this other, countervailing thing that we need it to match what existed already. Ultimately it was the latter group that won, so the Polarimetric data continues to be in this (older) NEXRAD message format.
>
> The legacy of those decisions continues to remain around, basically, even as we move on, we can't...There's a lot of people now set up using radar data with the old formats and the thing was do we drop all those old users in order to basically accommodate this larger diversity of users, or don't we? (PRI-35)

In this situation, we have a different incentive for influence on structure of data. Technical investments made by private sector users – external to the government system – are, of course, based on formats that exist at the time the investment is made. When government considers new standards, there is the pull they face related to their own systems to keep new data comparable with the old. But, on top of that there is the added consideration of the external inequity that could be caused by, say, moving to newer, more flexible structures for data that might improve data sharing, but would disenfranchise those organizations – commercial and otherwise - that depend on the old format and might incur costs or other issues if standards change. From this statement, it seems likely that the impact on external users of data are considered by the government in making decisions about formats and standards to use for the data. In Chapter 5, for example, we see NOAA providing expertise to one of the collaborators to help them embed the data in their tools to make it more usable. Whether this is specifically requested by industry or is done in anticipation of their needs and mutual dependence, may not be a critical aspect – the impact is there. Conversely, where government sees little benefit in the Restaurant Inspection case in Chapter 2, they are not willing to make this effort and a third-party steps in.

The Return of the CRADA: Private Sector Collaboration in Algorithm Development at the National Severe Storms Laboratory. There are a sizable number of public, academic, and private organizations involved in meteorological research, including NEXRAD. Lincoln Laboratory, for instance, is federally funded research lab for the Department of Defense at Massachusetts Institute of Technology and is involved in testing and development related to NEXRAD (MIT, 2019). The Center for Analysis and Prediction of Storms (as we shall see later in discussing access) is affiliated with the University of Oklahoma (University of Oklahoma, 2019), and the National Center for Atmospheric Research (NCAR) is federally funded, but run by a nonprofit consortium of universities (NCAR, 2019). Among other things, NCAR scientists have also performed NEXRAD algorithm research. However, the organization cited earlier by an official at the Radar Operations Center as being a source of ideas / research for NEXRAD algorithms is the National Severe Storms Laboratory (NSSL), a NOAA research laboratory. It is the NSSL, in fact, that developed a number of the original algorithms for NEXRAD (NSSL, 2019). Yet, there is an aspect of this organization that, like other areas of NOAA, involves collaboration with the private sector through use of a Cooperative Research and Development Agreement (CRADA) - the same type of agreement used by the NOAA Big Data Project. If the NSSL is a source of algorithms for the ROC, then their collaboration with the private sector might provide a conduit for influence.

Take for example a project from 2012 that established a CRADA between NSSL and Willis Re, a global weather reinsurance company, to conduct hail research. While the terms of these agreements require confidentiality, a press release at the time described the work and benefits to both parties. The organizations would "conduct collaborative research to improve estimates of hail size and coverage, allowing Willis Re to better quantify risks to people, property and agricultural interests" (NOAA, 2012). A summary of the benefits includes the planned impact of this work on NEXRAD algorithms:

> Both parties expect to benefit from the CRADA. Willis Re funds the Willis Research Network (WRN), which supports research to improve the quality of re-insurance and risk management of populations exposed to severe storms. *NSSL will be developing enhancements to the hail algorithms based on comparisons to hail insurance claims and Willis Re/WRN will be evaluating the performance of the algorithms* [emphasis added]. Willis Re and WRN will provide expertise in the development, use, delivery, and feedback of severe weather analysis modeling to the insurance industry. This collaboration is expected to improve the accuracy of insurance industry estimates of hail damage caused by severe storm events occurring throughout the United States and worldwide. (NOAA, 2012)

In a report from their most recent 5-year review, NSSL listed CRADAs that involved algorithms included hazardous weather algorithm development with Weather Decision Technologies and development, enhancement, and commercialization of severe storm algorithms with Weather

Services International, two private sector companies (NOAA, 2015). It is clear, then, that the NSSL collaborates with the private sector as a matter of course in at least some algorithm development. And, as explicitly stated in the excerpt from the press release above, the work is expected to have mutual benefit. Considered in retrospect, the CRADA is essentially a legitimized form of mutual influence in the form of collaboration that contemplates potentially mutual benefit. In the case of algorithms, these relationships are not pursued by operations (the ROC, as we saw above), but by the research arm of the program, where, of course, design occurs. The significance? When the applicability of this finding is considered more broadly, it seems likely that many data-producing functions of government may not have such an organizational analog for research. But, if we expand the definition of research to private and non-profit "think tanks" or advocacy organizations, it becomes apparent that there may well be a number of "labs" operating in a wide variety of program areas in government that could influence the design of information collection and even distribution as part of their policy recommendations.

Archive (Retention)

Retention of data is tightly connected to its value. Preserving observations beyond the time of initial measurement allows multiple value frames to be applied to it outside a real-time context. For example, historical observations can be used to validate the predictive or diagnostic power of a theory or – in the case of radar - the accuracy of an algorithm. In some cases, accessing data after it is gathered, retained, and processed is the only approach available. In this section, we look for evidence of attempted influence by the private sector on the retention period of NEXRAD data, along with influence on its content and/or structure.

When data reaches the archive, it will have already been defined by design and operation of the radar instrument and the application of algorithms. However, there is still the question of how the retention period is determined and how policies at the archive may relate to access and related distribution schemes, and whether or not the commercial use of the data affects these decisions. As we saw in Chapter 3, the NCEI (formerly NCDC) archive at NOAA is the key access point for external parties to environmental data and they are aggressively looking to market their holdings for a variety of policy purposes, including to the private sector. As in the phases of the information lifecycle previously examined, technology plays a part here as well and its evolution impacts both archiving and related access strategies that are, by necessity, shaped by decisions about the formats and technologies used to retain the data.

Origins of NEXRAD data retention decisions and policies. As noted earlier in this chapter, data from previous radar systems used by NOAA has been retained permanently at the archive. Prior to being able to remotely dial-in to radars to access radar information, access to data was provided solely from NOAA's archive at NCDC (later NCEI). Based on what we have seen so far, one might assume that industry would favor permanent retention of the entire inventory by NOAA. In this way, they could avoid the cost of maintaining the data over time while having unlimited access to it. So, the interests of the government and private sector here may be aligned. Thus, because radar data (whether it be in the form of data, products like images, or even photographs of radar screens) had been determined early on by NOAA to have value for research, it is possible that the question of industry's interest did not come up.

As context for the early decisions about retention of NEXRAD data (its volume was much more substantial that previous radar systems), it should be noted that NOAA was under scrutiny around this time on the question of how they were planning to care for and preserve the enormous set of data and records they were already attempting to store (GAO, 1990). This is relevant not only in that it may (see below) have provided an opportunity for industry to step in to maintain the inventory, but that it might have created internal pressure to save less – perhaps only what was most used/useful – or to make investments in archival technology that might also have external benefits (say, by moving pictures to be scanned and digitized for preservation, they might be easier for third parties to make use of and be easier to maintain). The situation is best evidenced by a General Accounting Office report from 1990 that evaluated the records management of both NOAA and the US Geological Survey (USGS). The report states that at the time NOAA had "440,000 magnetic tapes, 374 million film records, and 89 million paper records" (GAO, 1990) numbers which have been dwarfed by the rapid expansion of data collection via satellites in the last decade. It also notes that the data was irreplaceable and should be considered "a valuable national resource." The main findings relevant to our topic were that NOAA did not have an inventory of its holdings and that a low priority had been placed on managing and maintain the records (GAO, 1990).

Behind the latter observation lies another relevant finding. NOAA's data management challenge was, in fact, seen by Department of Commerce officials as an opportunity for industry to step in. The report found that the Department of Commerce had directed NOAA to conduct privatization reviews in support of its belief that data management and archiving should be outsourced to the private sector. However, while little privatization occurred, the findings stated that the results of that effort were "devastating." Budgets were cut "in anticipation of savings that never materialized" due to lack of private sector interest. And, internal initiatives to address the problem were put on hold in anticipation of this outsourcing (GAO, 1990). As an aside, it is interesting to note that, some 25 years later, NOAA would actively recruit industry hosting for distribution of data in the Big Data Project as a way to address their own resource challenges in hosting and distributing an even larger amount of data.

The GAO report proved a catalyst for new efforts by NOAA in this area. While NOAA had undertaken some prior analysis of the problem, they now began a concerted effort (cited in their response to the report) to establish an agency-wide data inventory, evaluate the condition and scientific value of the data in it, and work with the National Records and Archives Administration (NARA) "to develop NOAA records disposition guidelines" (GAO, 1990). From this, I take it that little industry influence was exercised on decisions about the retention period of radar data up to this point (which was essentially permanent to the degree data in the form of photos was captured from the older radar systems), nor for NEXRAD data, which, for Level 2 and 3 data is now retained permanently. But, the journey to successfully meeting that retention standard during the first decade of NEXRAD implementation was troubled. And, in those troubles lay the motivation (and justification) for later decisions about distribution and access addressed in the section analyzing that phase of the information lifecycle below. To set the table for that discussion, a brief summary of the initial archiving methods used for NEXRAD data and the associated problems follows.

Archiving NEXRAD data. As noted earlier, there is a tight connection between decisions made about how to retain data (especially the technology and format used) and the subsequent structure of design of distribution and access for secondary users. Simply put, for example, if documents are retained on paper in a file cabinet, then access is going to have to be in person or by mail unless a change in format or "technology" is made to enable other modes of access (scan, fax, email, etc.). In the same way, while NEXRAD data was streamed over network connections real-time to the weather field offices, decisions about the technologies used to record it for later analysis set the table for what access could be provided later to researchers – and the private sector - and how it would be designed.

As discussed earlier in this chapter, four levels of data are produced by NEXRAD weather radar: Level 1 data, the analog returns from the radar; Level 2 "moments," such as reflectivity, that are derived and sent on to the Radar Product Generator (RPG); Level 3, the set of "products" created by the RPG; and Level 4 which are custom creations using Level 3 data via the Principal User Processor (PUP) in weather field offices. In planning for the NEXRAD project, NOAA determined that they would eliminate the dial-in access previously provided to external parties as it imposed a burden on local offices (Baer, 1991) and replace it with a network distribution system run by commercial companies that was dedicated solely to Level 3 data (Level 2 data was not distributed to business or the public). This distribution approach (the NEXRAD Information Dissemination Service or NIDS) is described in further detail in the next section as part of analysis of the Access phase of the information lifecycle. However, actual recording of the data for delivery to the archive at NCDC was not addressed via this method.

As is still the case today, no provision was made for recording / archiving Level 1 data. In turn, the initial plan was to collect Level 2 data on a subset (Crum, Alberty & Burgess, 1993) of the radars using portable recorders that could be plugged in to a slot on the RPG. It was understood that the Radar Operations Center would coordinate the decision on which radars would receive the recorders. Level 2 data was recorded to reusable 6mm tape and was accompanied by a carousel that would cycle through automatically loading 10 tapes. Level 3 products were recorded to optical discs that were not reusable. Finally, Level 4 data was recorded at user discretion using the same technology as Level 3. At initial implementation, there were no plans to collect Level 1 data centrally. While Level 2 data was collected only at selected sites – and was voluminous – what was recorded would be sent to the archive at NCDC; Level 3 data was to be collected centrally and was considered the permanent NEXRAD data record. Level 4 was considered to be for local use (Crum, Alberty & Burgess, 1993).

Deploying the new NEXRAD radars across the nation was a multi-year project. As it was occurring, there was increasing awareness of the value of this data for research purposes, and the gaps that were left by recording that was only at select sites and not continuous (NEXRAD TAC, 1994). Consequently, there were internal discussions about the need to extend Level 2 archiving to all of the radars even while the limited approach was rolled out. After approving a Level 2 data collection plan in June of 1993, by September, after hundreds of the tapes from the portable technology were recorded and archived temporarily at the National Climate Data Center (NCDC – the precursor to NCEI discussed in chapter 3), the NEXRAD Configuration Control Board approved a change request to install the recorders at all NEXRAD sites (NEXRAD TAC, 1993). The understanding, then, about the research value of Level 2 data was evolving, along with decisions about its retention. Once the equipment and archiving approach was in place, they settled upon a routine where Level 3 discs would be sent to NCDC as would the Level 2 tapes. However, the tapes were reusable, so the data would be copied from the tapes before sending them back, going through a series of transformations to be archived and, over time, were written to a disk array at the NCDC. Figure 30 below shows the process for Level 2 data as of 2000 when they began transferring the data on the tapes to a disc system:

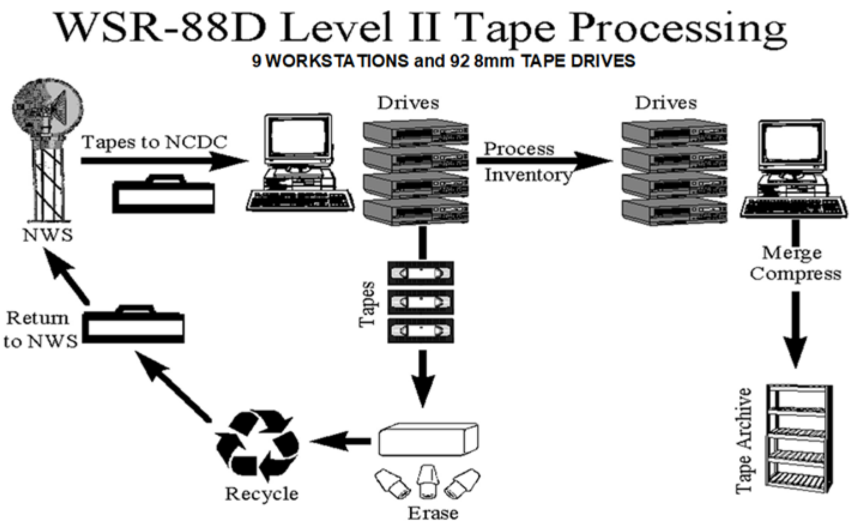


Figure 30 - Level 2 Archiving Approach (Del Greco, 2003)

The result, however, was far from perfect. There were limitations in the reliability of the tape technology and, because of the storage methods used, significant constraints on later access. An article about a subsequent (Droegemeier, et. al., 2000) archiving and distribution system discussed in the next section summarized the challenges created by this approach:

> Neither the recording media nor the hardware were designed for the stresses to which they are subjected in the field. Consequently, the tape system failure rate is high and partially explains the low archive rate noted above (Note: By 1999, the Level 2 archive was only 64.9% complete). The slow speed of the 8 mm tape drives, 92 of which are presently maintained at the NCDC, lengthens significantly the time needed to both read and write Level II data. Consequently, an order for even a modestly large data set involving a dozen radars may take a few months to be processed, and the cost for the data, especially for use in long-term, large-area climatological studies, can run into the hundreds of thousands of dollars -- far more than can be accommodated by a typical research grant. (Droegemeier, et. al., 2000)

The Level 3 archiving method was more effective, with about 90% (Droegemeier, et. al., 2000) of the archive data complete as of the same time – however, the disc technology eventually became obsolete. As we shall see in the following section on Access, these deficiencies were used as part of the justification for changes in the distribution of data to internal and secondary users that addressed these problems, too.

While, for NEXRAD data at least, we do not see influence from the private sector in its retention, this exploration has not been without value we have established this phase as a site of possible influence. Decisions made about what data to keep and for how long, and the technologies and technical approach used to retain it have the potential to directly affect the availability of this resource for commercial use – both because it determines if the data will exist and because those decisions are tied in to what access methods can be used. We see this explicitly in our discussions in the next section about Access – most recently in the Big Data Project, where the legacy of decisions about how NEXRAD data would be archived and managed resulted in difficulty in commercial access that provided a rationale for moving it to the cloud hosting of the collaborators during the Big Data Project in Chapter 5. In turn, we see that decisions and changes in technology that occur in different places in the lifecycle (moving from photographs of radar screens to electronic products, for example) are interrelated and that planning and design in each phase can be a place where intervention continuously presents itself as an option to improve private sector access and opportunities.

Access (Dissemination)

Technology, Path Dependence and the Evolution of Radar Data Dissemination. In the previous section, we saw NOAA arrive at an archiving approach for both Level 2 and Level 3 data, with both types residing at the National Climatic Data Center (now the National Center for Environmental Information or NCEI). From the previous iteration of radar, NOAA had knowledge that the private sector was making use of the Level 3 data, to the degree that in some cases there were as many as 100 lines running directly into a NWS office (Baer, 1991). With the transition to NEXRAD, direct access by external parties would be eliminated. We take up the story with the genesis of the methods used to allow NEXRAD access and consider what role, if any industry played in the relevant decisions.

NEXRAD Information Dissemination Service (NIDS). NIDS was the initial approach implemented by NOAA to address access by the private sector. It differed significantly from methods used in the past. With NIDS, each radar site would have a four port switch that allowed four identical outbound streams of information from the radar to be made available to the private sector. However, instead of individual companies attaching to them, access to these devices was granted under reimbursable memorandums of agreement with three private sector providers: Alden Electronics, Kavouras, Inc. and WSI Corporation (precursor to The Weather Company) through September 1999, who, in turn, agreed to further distribute the data to the private sector (Baer, 1991; Gratz, 2005) (see the box in the upper right of Figure 31 below labeled NEXRAD Information Dissemination Services (NIDS) Product Suite (4 Ports)).

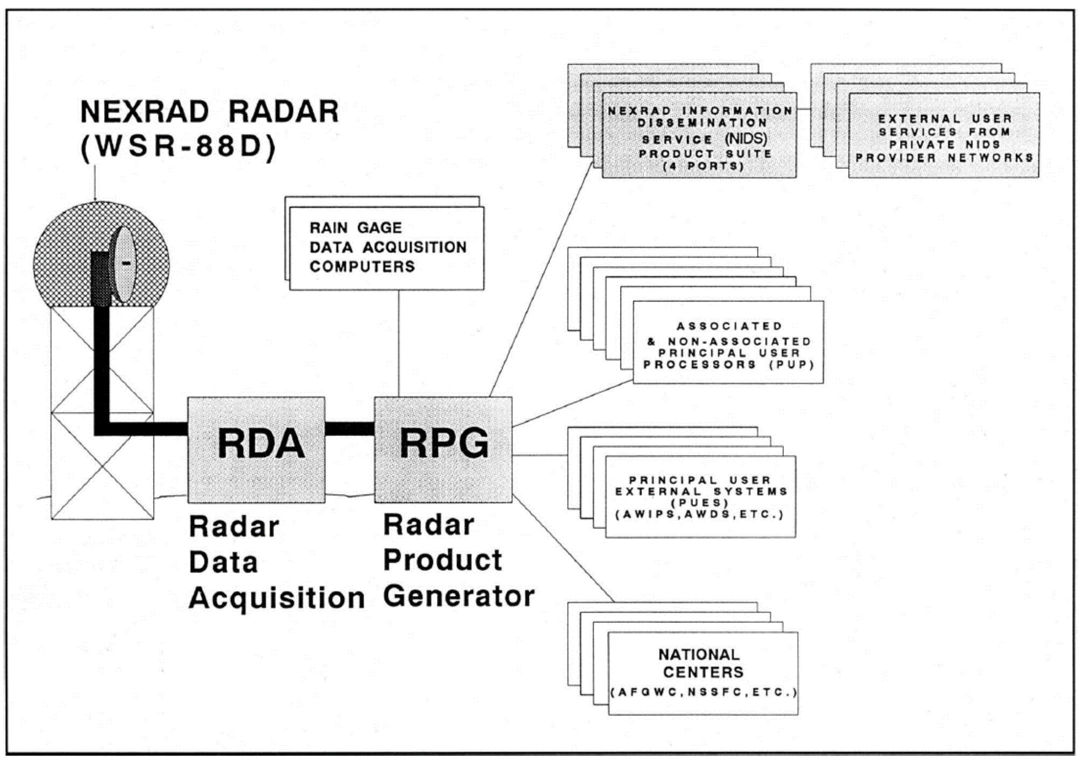


Figure 31 - Radar Product Distribution (Baer, 1991)

This change in the approach - providing access through agreements with private sector companies to distribute NEXRAD data - came about through the collision of several forces: 1) A change in federal policy that allowed an arrangement like this. Under the Omnibus Budget Reconciliation Act of 1990, NOAA could charge the “fair market value” for data (15 U.S. Code § 1534). While NOAA chose to operate under a cost-recovery model (vs. requiring they be paid fair market value for the data), they decided to grant a franchise to intermediaries to resell it in the open market; 2) Improvements in computing and networking technology that allowed Level 3 products to be distributed electronically in a reliable and timely manner; and 3) An existing demand for commercial (external) access that had become a requirement. The idea of allowing the three private sector companies to serve as brokers was justified by the National Weather Service under the theory that these brokers would have natural incentives to compete to provide the data to commercial customers at the lowest cost (Baer, 1991). The providers were selected via competitive bidding and they were charged only the hardware and maintenance costs incurred by NOAA to grant access to the data (cost-recovery, as noted above). The data providers were required to make data available unaltered from “any and all” current and future NEXRAD sites (Baer, 1991). For non-profit and other users who had previously been granted access for no charge, a coupon system was set up that would allow them to be redeemed with the data providers to continue this subsidy and avoid fees.

In this case, then, while the providers were allowed to make a profit on the service, the NWS was actually a) not paying them anything (as in a normal contract) and b) only receiving reimbursement of the marginal costs of distribution in return. This is an interesting contrast to the model in the Big Data Project where distribution by the companies is required to be free, but there is an assumption the collaborators (companies) will make money from those users who choose to process the data on the platform in place rather than downloading it – a concept that was not feasible technically when NEXRAD was implemented. And, as we've seen, has yet to be proven financially feasible now. In talking with executives involved with this distribution method at the time from both the government and a company that distributed the data, it does not appear that these companies exercised special influence to obtain the franchise (PRI-36, PRI-37) - and it was publicly bid. In fact, the cost to the data providers over time was significant (between $250,000-$500,000 annually - Gratz, 2005) – see the comment below on their revenue.

The capacity of networking technology available to NOAA was also growing, along with demands within the National Weather Service (NWS) for integration of data sources beyond NEXRAD (hydrological, satellite, for example) for use in weather monitoring and prediction. Toward the end of the decade, the NWS built technical infrastructure in support of a new integrated Advanced Weather Interactive Processing System (AWIPS) that combined access for many types of weather-related data into one interface for use by the weather service (Raytheon, 2019). This infrastructure also provided a method for the NWS to collect and make Level 3 data available directly to external organizations without the need for the data providers contracted with through NIDS. By 2000, a new service called *the Radar Product Central Collection Dissemination Service* (RPCCDS) began collecting Level 3 data centrally from all the weather radars (Gratz, 2005) and making it available to commercial parties. At the same time, the NIDS agreements with the data providers were terminated. The RPCCDS is still in operation today, although users also have other methods available to obtain NEXRAD data through NOAA. The general understanding of the situation of the NIDS data providers on the NOAA side was "they didn't make any money" (PRI-36), and the move to free Level 3 data from NOAA was positively received by the private sector. While the Level 3 data continued to be archived on optical disk during the time NIDS was in place, shortly after they took over distribution of the data, NOAA began sending it directly to the National Climate Data Center using this network and eliminating the optical disc recording method.

What can we learn here? As we move through the last stages of this discussion, we can see that changes in technology make wider distribution within NOAA first possible, then cost-effective and, in turn, are then leveraged to make information available to the private sector. NOAA is aware of this external demand and, even if it is not explicitly included in their decisions about the data itself, they understand the need to provide access to the data and the dependence of the private sector on it.

Collaborative Radar Acquisition Field Test (CRAFT). This project was the next significant step in the evolution of NEXRAD data dissemination. This development focused on Level 2 data, the higher resolution base data that is processed further by algorithms to create Level 3 products. While networking and computing technology were not sufficiently robust and too costly to distribute this volume of real-time data widely when NEXRAD was first implemented, they continued to improve. As we saw in the earlier discussion of the Archive phase, the value of this unprocessed data was seen as significant enough to create a tape archive, but this was the extent of the effort at that time. Then, in the late 1990's, the University of Oklahoma's Center for Analysis and Prediction of Storms (CAPS) and others developing numerical models of storms conducted research that showed the significant value that could be added to storm prediction and analysis through use of Level 2 data (Kelleher, et. al, 2007). This, coupled with growing interest in access to the data within NOAA, in the academic community, and in the private sector – and advances in computing and networking technology – made the time right for a change. An executive at CAPS at the time explains:

> [W]e were working with a lot of private companies in this center, this Center for Analysis and Prediction of Storms, called CAPS. That was the name of the science and technology center. Plus, I look at this thing, I said, "Wait a second. The National Weather Service is in what department of the federal government?" The Department of...Commerce! Oh, my God. It's in Commerce! Okay, well, you know, a lot of other companies were telling us if we could get access to the Level 2 data ourselves, boy, we could create all kinds of hydrologic products, and all this stuff. We can't do it with NIDS, because there's no way to customize NIDS. NIDS is a product (Note: It distributed Level 3, already processed, data). If every company gets the same product, you're basically stuck with whatever NIDS is. It became very clear that there was a compelling need for Level 2 data, far beyond just our own narrow interests. (PRI-38)

Similar to what we saw earlier with the NSSL's work on algorithms via CRADAs, academia again provided a gateway to collaboration with the private sector not available to the operational part of NOAA:

> This is why OU (University of Oklahoma) is so important in CRAFT, is because the weather service and NOAA gets very nervous, and most government agencies get nervous, is they're talking to company X, Y, or Z. It's like, "If you talk to X, Y, or Z, you can talk to A, B, C, D, E, F, and G, too." We said, "No. As a university, we could talk to whoever we want to." People were very comfortable working with us, because there wasn't that government-ness thing to it. We're this neutral ground. (PRI-38)

The latency of the delivery of the data (how close to real-time it could be delivered) was the central problem to be overcome. Progress in this endeavor came into sight when the NSSL deployed T-1 connections (1.5MB/sec) that allowed direct access to Level 2 data from NEXRAD radars on an experimental basis with no negative impacts (Jain & Rhue, 1995). They called this system the *Radar Interface and Data Distribution System (RIDDS)*. As technical feasibility and increased demand converged, a consortium of government, academic, and eventually private sector parties came together in 1998 to form the *Collaborative Radar Acquisition Field Test (CRAFT)* project to investigate the possibility of wider, real-time distribution of Level 2 data (Kelleher, 2007).

The success of the CRAFT project depended on several advancements in technology: The RIDDS architecture that introduced a higher capacity connection directly to the radar, then, further upstream, access to high speed national academic internet "backbones" (Internet 2 and Abilene) for wider distribution; and a software product developed by the University Corporation for Academic Research (UCAR) called Local Data Manager (Unidata, 2019) to manage the distribution of the data across nodes on the computer network. The size of the data collected could also be a barrier to its delivery in near real-time. Here, the answer came in the form of a compression algorithm (bzip2) that made it possible to use a 128 kbps connection (quite slow by today's standards) to transport the data with latency of 60 seconds or less to NOAA's National Center for Environmental Prediction (NCEP) with actual times reaching as low as 10 seconds or less (see below) (Kelleher, et. al, 2007). To re-center the discussion to today, the reader might recall that it was high-speed access by the private sector to data held at the NCEP that was the purpose of the ill-fated Reston Project addressed in Chapter 3 and 4. So, like today, here we see that access to data - and, thus, the opportunities for influence - are also tied up in "the possible," combining progress in different technologies – in this case, not just the radar itself, but networking and compression technologies – that involves planning, investment, and testing, where decisions are spread over time and involve multiple stakeholders. It may be, then, that across a variety of government programs, data may lack value (and thus fail to attract dependence and influence) simply because of its size and because cost-effective access is not yet commercially feasible – see, for example, NEXRAD Level 1 data today that is not universally archived. The private sector algorithm developer we heard from in the Algorithm (Processing) section conjectures on the possibilities:

> The thing to think of, you know you get into Level-1 data and you're literally looking at raw voltage data essentially, and phase data. And what that would open a private sector company up to do, or anybody really, it doesn't really matter...You could start to process the data your own custom way. You could say "I'm not going to use a square window when I run across here to process this spectrum of data. I'm going to use a Gaussian

> window. I'm going to do oversampling here." And you can build your raw moments, that Level-2 data-you can build those moments in a different way. (PRI-34)

In this case then, he is talking about accessing the date before it reaches "programmable signal processor" we heard about before ("raw moments"). Today, that data is not archived, nor available beyond periodic recording at the radar location used for research projects. Does the story sound familiar? In that case, the future of distribution and use is still cloudy. For now, we conclude by examining the final steps that made Level 2 data available to the private sector.

Collaboration and Coordination with External Parties. In numerous presentations, journal articles and documents about the genesis and implementation of the CRAFT project, private sector interests and participation are continually acknowledged. The project itself was made up of participants from the academic and government research communities, along with representatives from what is now the Radar Operations Center. As the technical details were being worked out and various scenarios and technologies tested, they conducted a series of workshops for the overall weather community to present what was being learned and get a better understanding of the feasibility (financial, technical) of various options for the distribution of Level-2 data, along with industry requirements. These workshops were all well-attended by the private sector.

In the first workshop, held in 2001 (University of Oklahoma, 2001), the private sector attendees were given time to "summarize their perspectives on requirements for (access to) Level 2 data." Following the workshop, they were granted access to a selection of participating sites to conduct testing. The stated purpose of the next workshop (2002) was "[T]o establish the needs and requirements of non-NWS users of WSR-88D real-time Level 2 data, and to define a framework for meeting these needs in both the short and long-term" (University of Oklahoma, 2002). It included presentations from companies on their use of NEXRAD data. The outcome was private sector consensus on the need for 99.99% uptime and 10-second latency in delivery. An 11-member committee was then formed, with five industry representatives, to complete the work. Around the same time, CAPS reached out to the private sector to obtain additional resources to continue the project, issuing a "Call for Private Sector Participation in the Collaborative Radar Acquisition Field Test" that sought to "encourage the continued development by private companies of products and services using the present CRAFT data stream, which involves 59 radars, with costs shared equally among all private sector participants" (University of Oklahoma, 2003).

The final approach was presented in a 2003 workshop, with 17 private sector companies represented (NWS ROC, 2003), along with a presentation by NOAA of eight possible access models for consideration. In addition, NOAA indicated it could only commit to 95% availability. In the public comments that were solicited following the meeting, there were two interesting reactions. The first was a shared belief that using the private sector to broker / deliver the data like it had in NIDS was not acceptable ("private gatekeepers") and would end up with the larger providers, the only ones with the resources to do it, subsidizing smaller companies (NWS ROC, 2003a). This was the reason professed, but remember the lesson from Reston and the threat of privileged access by the company delivering the data that might tip the scales in its favor. In turn, remember that the companies had likely already found that there was no money in it. The second observation also anticipated the Reston project model discussed in Chapter 3 and 4 – anticipated the Big Data Project model as well, asserting that "[E]xisting data dissemination paradigms" were "ever more challenging and cost-prohibitive" suggesting that "future dissemination services should provide accommodations for access to and placement of vendor owned and operated processing capabilities as far upstream as possible in order to reduce the data volumes and optimize communications requirements" (NWS ROC, 2003a). So, in this case, we see industry, six years before the convening of 2009's Environmental Information Services Working Group (see Chapter 3), already lobbying for an approach that would allow them to move their "compute" closer to the source of the data, rather than having the government invest in distribution networks that allowed access, but introduced latency and would potentially fall short of industry needs in the future.

Meeting the Mission: Improving the Quality of the NEXRAD Archive as Justification for Private Sector Distribution Improvements. In these discussions, the leaders of the CRAFT project were well aware of the problems highlighted in the last section with the quality of the Level 2 archive. As stated matter-of-factly in an article about CRAFT co-authored by the project principals (Kelleher, et. al, 2007):

> The NWS allocates resources based upon requirements. Therefore, it had to be demonstrated that CRAFT technology would help NWS meet its requirements before investments would be made by the NWS to make CRAFT-like technology operational. NWS had an ongoing requirement to archive Level-2 data. However, it was clear by 2004 (Note: the CRAFT project began in 1998) that the 8-mm tape recording system in use at that time would become logistically unsupportable.

In talking with a leader of the CRAFT project, it appeared that the work here was to gain final commitment from the NWS to the new distribution model, thus getting the data to academia to fuel research and delivering it to the private sector who could use the same method. Chaining the need to archival requirements was a little more strategic:

> Well, the weather service wasn't terribly interested in that, because they exist to meet their own mission requirement. Their mission requirements were being met, because every NEXRAD had a PUP, it's Principal User Processor, and it was getting Level 2 data. We worked hard to convince the weather service that, "Hey. You're in the Department of Commerce. There's a compelling need for this for research, and the Abilene Network exists, and so on.
>
> At that time, in particular, there was a frayed relationship between the private sector and the weather service. They were not at war, but they were at odds with regard to who

> should provide what. There was an academic study done called *Fair Weather* (see Chapter 3), that got into that. We felt that, as a university, we provided that neutral brokering capability, and we said "Look. We could show you that we need the data for research, and the private sector needs it. How about we, meaning the University of Oklahoma, how about if we make this data available to the private companies provisionally?"
>
> So, as part of selling the project, we sold another benefit, that of replacing the burdensome tape archive project with a real-time feed to NCDC to increase the quality and timeliness of the data. And, once CRAFT became operational, this approach was implemented. The win was for the NCDC (the archive), the win was for the private sector. They got access. They developed products, and services, and stuff. We got the data for research, and everybody was happy. (PRI-38

In Level 2 data, then, we have an object case with which to look at industry influence. What we see, however, appears to be an entrepreneur inside academia in need of real-time access to data for research bringing together multiple interests – including those of the private sector – to introduce a new approach to access / distribution. Industry was definitely consulted on their needs. In fact, at the last workshop where the distribution proposal was presented, the notes from the meeting stated that the head of the National Weather Service "[E]mphasized that the network is being implemented to meet NWS requirements, but where changes can be made at little or no additional cost to accommodate external need, the NWS will definitely consider them" (NWS ROC, 2003b). This story reinforces the idea that the agency is well aware of private sector interests and requirements in their design of access and distribution and considers options where it can support them – when there are ways they can also benefit.

Conclusion

When we consider the research questions that underlie this study - the extent to which the private sector seeks influence, how they exercise it, and the impacts it has on government – it is hard to miss the role that technology plays in the answers found in this chapter. It is information technology after all, so it is natural that its capabilities, and limitations, would affect how and if dependence on secondary use of government information develops and its impacts. By using an information lifecycle approach to the long history of a major government computer system that produces data of great commercial value, we have also seen where opportunities for construction of information lie and the interrelationships between choices across each phase that impact the answers as well. Bringing these two threads together creates a lens for what we have seen in the study thus far.

Weather radar has a long history, one where commercial uses were identified almost immediately, but the evolutions of its capabilities and private sector ability to use them has been dependent on technology. Dependence grew, but adequate delivery methods grew more slowly. And, when a new system was designed, the focus was on government's needs. While a huge investment with plenty of private sector involvement, when it was deployed NOAA sought to minimize the government's role in access to the data and the "overhead" that came with it – offloading it to the private sector, not for profit but for convenience, while still learning itself about the potential value of retaining the data for its own research. The benefits to government here, even though NOAA calculates them to be dramatic for society as a whole, don't seem to increase their interest in benefitting from the private sector's dependence. For example, it was only once a technical solution became feasible for meeting their retention needs for Level 3 data that they allowed the private sector to use it, first with Level 3 data and only much later, with the prompting of an entrepreneurial academic center demonstrating its feasibility, did they see their way clear to do essentially the same thing for Level 2 data as they had done for Level 3, allowing the private sector to access it using the same technology that fed their archive. While we look hard for influence in algorithms, the code that creates the products so valued by scientists and the private sector, we see what suddenly becomes obvious about open government data – the transformation and value-add in the form of improvement or customization is where the private sector makes their money. Like their efforts to restrain government competition, the data the private sector uses only needs to be good enough for them to add value – any subsequent improvements accrue not just to NOAA, but to their competitors and even the public. The words of the loaned executive in Chapter 5 ring clear – and remember, this was an attempt to make this same data, NEXRAD even more widely available through the Big Data Project: “And it was wild having been at (one of these companies) and using the data, and then coming out and them telling me it wasn't useful. I was just like, “I don't know what you're talking about. It was useful when I was there” (PRI-22). And so it was, but as he continued “They thought “Well, that's just going to help other people, not us.”

In a twist, this seems to be the same logic that powers government motivations here, and really a main lesson of this study. Where government can benefit, or thinks that they might, they are willing to put some skin in the game, potentially changing their practices. Where they can’t or don’t think they can, they ultimately default back to their mission and the prohibitions imposed by their resource and technical limitations, a legitimate concern. This leaves the private sector to find and convince government of some benefit, or at least find a way to minimize government’s cost and inconvenience. All this is set in a context of equal access for vendors, but unequal capabilities, technological and market change, and planning that recognizes the dependence of private uses, but puts its focus on accomplishing the mission for which the agency was designed, staffed and funded.

Chapter 7: Discussion, Findings, and Conclusion

Freedom of access to government information is now part of the bedrock of American democracy. It would be hard to conceive of any policy with a stronger "ideational and symbolic legacy" (Beland, 2010) than one that grants citizens the right to know what their government is doing and has done, and with this information, to hold it accountable. However, as it has in almost every area of society, developments in the power and use of technology have, of their own accord, expanded the forms, uses, and value of information – and government is no exception. Sometimes these come to public awareness through services like phone alerts for missing children or online interactions with the IRS. But, for administrators with exposure to the internal operation of government, information systems are everywhere, in every facet of what they do and the way programs are managed, delivered, and assessed. These new capabilities have also expanded government's collection, processing and use of information to facilitate and support the myriad of their internal and external functions. In both cases, the collection and use of government information has become deeply interwoven in the day-to-day life of administration and society.

In the small corner of the world touched on in this research, we witness these forces converge. What once were policies intended for holding officials accountable for their actions now provide a gateway that allows information byproducts of the machinery of government to spawn and power giant private industries, attracting their interests and resources to the table in matters that, for the mission of government, had previously been only a secondary concern. We find, however, that while this power comes with demands, it can also be leveraged by agencies to achieve benefits of their own, ones that, as time goes on, they come to depend on to maintain, or even achieve, their mission. And at levels of government where there is no prohibition on selling information, government may develop even more incentive to pay attention to such interests as their programs come to rely on them to fund some or all of their execution. In all these cases, however, information is a commodity and government is participating in some form in a market for it, with all the attendant risks of substitutes, and changes in equilibrium as the private sector seeks to maximize their profit. Friends today, and forgotten tomorrow.

In this study, we have heard the people involved in these relationships speak, both public and private, sometimes at (too much) length, about their interests, their opinions, and the way they make decisions related to this relatively recent phenomena. And, in ways big and small, we have seen some of its current – and potential - effects. But during these conversations, one might notice that as much as they were dominated by this legacy of openness, equity, and opportunity associated with freedom of information, the voices seemed to be, for the most part, winging it. There are business strategies and tactics, but for government actors, it was mostly exploring, reactive, without much in the way of examples or references from other domains, best practices, policy guidance - or oversight - to direct their work.

From this viewpoint, they are operating in a policy vacuum, specifically as it relates to identifying and managing the impacts and externalities of these relationships. But, when we see one developed at NOAA to address the provision of public data to the private sector, it serves to embed more private sector interests and requirements, carving out only the most basic elements of primacy for the government's mission. For you see, business is a public after all, and these freedom of information policies must apply for them. Dependence cannot be stopped by adding fees or forms, nor can government be restricted in granting access to information, the public interest is too strong, as is the need for responsiveness – to all publics. So there is no easy solution to managing the effects of dependence. Well, except one: To subject these influences and mechanisms and impacts to the same disinfectant that brought these policies to life to begin with – Sunlight (Brandeis, 1913). Further research, rigorous reporting, and education so that government managers and the public itself can see and study this phenomenon of mutual dependence for what it is and what it does, both good and bad, its risks, and rewards. Such transparency about this subject is only fitting to be handed to the people in a democratic polity, where information, too, is power.

Across the preceding five chapters, I have examined seven cases of involving aspects of reuse of government data by the private sector:

- Four of the cases involved specific types of data. In one, state government received revenue from the sale of driver history records to industry that were not open to the general public. In the second, state government received services in exchange for grant of exclusive rights to industry to resell access to genealogical data that in its native form, paper, was an open record. In the third, state and local government received nothing in return for use of data they produce and make available free online, while a private sector company turned to work with another company to obtain and transform the data it used, but added little to the bottom line. And, in the final case, weather radar data is made available free of charge to industry, potentially benefiting government by avoiding costs of delivering some services that the private sector uses the data to deliver.
- Three of the cases involved policy and programs within the National Oceanic and Atmospheric Administration (NOAA) that relate to private sector use of data. The first examined the influence of industry on a government policy for providing environmental data to the private sector, while the second looked at a program associated with their large archive of environmental data that attempted to promote its value to industry to gain 1) support for

NOAA environmental programs by demonstrating the value of the data they create and 2) to incentivize the development of a market for climate adaptation services in the United States. The final case, spanning two chapters, examined the development and implementation of a project undertaken to devise a business model for industry that would allow them to host and distribute NOAA data free of charge to the public (with NOAA avoiding that cost), while subsidizing that service by an initially unspecified "value-added" activity. The data involved in all three of these cases was open and available without restriction to business and the public.

While each case study included excerpts of documents and interviews germane to its subject and policy area, comments were also made throughout on the relevance of examples to my research design and questions, summarized in the conclusion to each chapter. The following section presents an overview of the findings for each of my research questions, followed by a discussion of the areas that are the most significant, with implications for theory and practice. The final section presents my general conclusions along with several recommendations for further research

Research Findings

My findings are presented below in terms of the research problem and questions laid out in the Introduction to this study. The problem statement applied policy feedback theory to the relationship between government and private sector organizations that depend on its data to produce revenue. This theory suggests that in such cases the private sector would attempt to influence government policies and practices to ensure a continuation of these benefits, as well as to maintain or increase their value. Thus, the resulting research questions explore the mechanics of this phenomena to understand first, if it exists and, second, if so, some of the conditions that may serve to block or enhance its effects.

The first question was "*What private parties have developed an interest in governmentally held data?*" While this study by no means claims to be a comprehensive survey of instances of private sector dependence on government data, the private parties identified include:

- Driver History Records. Insurance companies, information brokers, intermediaries contracted by states to sell records to them, and many downstream users for background checks or other uses allowed by law.
- Genealogical Records. A private sector genealogy service catering to publics internationally (Ancestry.com)
- Restaurant Inspections. A major private provider of online business directory services and multiple companies that crawl state government data to sell to them and to other businesses such as insurance companies or restaurant franchise owners.
- NOAA environmental policy. Primarily large companies offering meteorological services as their main business with other businesses potentially using environmental data not identified in the case.
- NCEI data promotion program. As one purpose of this program was to identify success stories of users of their data, this list is long and includes companies in the following lines of business or industries: aviation, agriculture, reinsurance, retail and manufacturing, the power sector, livestock, logistics & transportation, fisheries, and weather service providers.
- NEXRAD weather radar data. This data is also used by some of these same industries addressed above in the NCEI data promotion program, primarily via weather service providers. Bur an interview was also conducted with an individual who developed precision hail forecasts for an international insurance conglomerate.

- The NOAA Big Data Project. This project attempted to interest several of the world's largest companies with multiple lines of business in this data, and, at least for the period of the project, Google, Microsoft, Amazon, and IBM (who purchased and absorbed The Weather Company, one of the two largest weather service providers in the United States) agreed to participate in researching commercial uses of this data.

In sum, across just several policy areas, there is widespread interest by the private sector in access to and use of various data held by government at the federal, state, and local level.

While it was established in these cases that these parties were interested in the data, the next research questions ask "*To what extent do these parties seek to, and gain, influence on the content, structure, or availability of these data?*" and, if so "*How do they exercise this influence?*" Finally, I ask "*What are the effects of this influence on the form or nature of governmentally-held data, policies on data availability, and decision-making based on these data?*" This question was answered by evidence in the form of interviews, policies and other documents presented in each case study and in its conclusion. Summarized again briefly for each case:

- **Driver History Records**. I found no influence on the content of these records, nor their structure. I found their availability was restricted by both federal and state policy covering the permitted uses for the data, and was not available to those who would use it for other reasons. I found evidence that industry successfully lobbied the legislature to change that law in a state to allow more uses, providing inaccurate testimony that was not objected to by legislative research or the state agency about the impact of those changes as expanding uses of the data. The influence I found here also occurred in several ways that were not related to availability or formal policy, but to preserving or expanding the revenue to the state, something that assumedly could also come from the previously described intervention. This included state coordination with other parties that participated in the use of the revenue from the data to price the data in a way that would maximize revenue in the market and minimize incentive to use in-state substitutes for the data, such as court records. The state also raised fees on records to obtain more revenue for itself, and in response to lobbying by an intermediary company whose work they benefited from in other areas (web development and other subsidized services) to help subsidize them. A government and intermediary worked together to block legislation that would have made a lower cost substitute record, the consumer credit report, legal to use for insurance ratings in order to prevent loss of revenue. Finally, in one example, the legislature overrode existing law via proviso on an annual basis to allow revenue to be used for other purposes than was provided for in statute.

- **Genealogical Records**. Here, the content of the records was not changed, nor their structure, as the activity involved digitization of paper documents. Their availability was restricted in the new electronic form via a contract with a vendor that allowed the vendor exclusive rights for a period of time to sell the digital copies. In practice, one interview indicated it may not be feasible for the electronic versions to be delivered any other way due to inadequate state infrastructure to do so, even though that exclusive right lapses. The policy impact here was that, while the state has an open records policy that applies to the paper copies of these records, the agency made the determination that the electronic copes were supplemental and thus the access to the actual (paper) records was not diminished, using this rationale to justify their decision to grant exclusive access rights to the private sector.
- **Restaurant Inspections**. In the case of restaurant inspections, the content of the records was not changed, nor the structure – a company's attempt to influence government to change that

structure essentially failed, as government resisted participation. The company attempted to exercise influence by lobbying and contacting governments to participate, and, in one case, using the press to pressure a government to drop a new policy that they felt would impair the value of the data. Availability of the data did expand, in that, while it was already available online at individual websites of government jurisdictions, industry paid another company to write programs to strip this information from these websites, consolidate it, and deliver it to them in the form the company had proposed for governments, where it was included on their high-traffic website and smartphone app. There were no policy impacts found as part of this case. However, decision-making based on this data was affected, to the degree that users of the data access it through the version made available by the private company, as in many cases it reflected the application of an algorithm to standardize the scores based on history and other factors, something the restaurant inspections do not do, as they are point in time.

- **NOAA Environmental Data Policy**. This case did not involve a specific data set. But, it did involve the development of a policy for government sharing of environmental data with the private sector that contained provisions addressing aspects of the content, structure, and availability of the data, among others (notifications of anticipated changes and a chance for private sector input, for example). The method by which the influence was exercised included working with a U.S. senator to introduce legislation to threaten to abolish key functions of the agency unless policy concessions were made, and by participating in committees that had been set up to provide oversight and input into NOAA operations as it related to data sharing with the private sector. Another technique used was to frame industry and government goals as similar, that they were all trying to advance weather science and help people and business. Industry created a sense of urgency about the changes it requested in policy or practice, and also worked to include provisions that left boundaries between what was appropriate for NOAA to do and the private sector to do indefinite, providing opportunity for future influence. It is not clear how decision-making based on this data would have changed as the case does not deal with a specific dataset or its use.

- **NCEI Data Promotion Program**. This case also did not involve a specific data set. However, a program that solicited input (feedback) from the private sector on changes or new requirements they had for the content and/or structure of existing data "products" did occur. The mechanism was through a program funded and worked on cooperatively by government. It also had a goal of expanding use of data, especially by companies who sought to provide climate adaptation services and developed a strategy to market the data by industry classification matched with their agency's statutory mission and program goals. While this case, again, was a study of elements of a program and their potential points of intersection with other parts of NOAA that could impact content and structure of data over time, there was an example found where NOAA changed the frequency of data provided in a product, which would allow for different uses and decision making.
- **NEXRAD Weather Radar Data**. There multiple sets of data produced by government weather radars. While this study did not find evidence of influence on the content or structure of those data, it did find evidence that both government and the private sector were interested in increasing the availability of the data and traces how that evolved over time. So, availability was subject to mutual influence. At first, the government set up a mechanism to provide the data on a cost recovery basis to several vendors who, in turn, sold it to other companies, which limited access – even though the data was technically "open." Then, the government took the data provision "in house," and eventually used expressions of industry

demand and their own archival needs as a reason to institute requirements that justified installing infrastructure that could meet both needs. Decision-making based on this data did not appear to change.

- **The NOAA Big Data Project**. This project focused on NOAA data writ large and proposed to expand the availability of data. Initially, neither government nor industry impacted the content or structure of the data. And, most of the types of data that were chosen to be deployed in the project were already available. Over time, interviews demonstrated that industry could obtain more revenue and use of the data if they transformed its format to deliver in their own tools and NOAA helped with that. This case addresses a project, not a specific type of data. While there were no policy changes coming from this project, the premise was a change in practice under the policy, by providing data via private sector platforms for no charge. The final contracts for the mechanism that was decided on going forward are just now becoming available and it is not clear yet if policy may change. As a specific type of data was not involved, and the data that has been deployed so far is identical, it appears no decision making based on the data changed in this case.

Discussion and Synthesis of Principal Findings

Throughout the conversations I have had in performing this research, and given my personal background in government, I have been struck by the general candor of not just government employees but even those private sector employees and executives I spoke with, and I think this candor should be visible to the reader in some of the verbatim conversations I recount across these chapters. What emerges for me is a very real situation – all my conversations have been about secondary – not primary – use of government information by the private sector. Those conversations, to me, seem unusual in the daily study and execution of government business and, given that there is no central inventory of these relationships at any level of government, it is hard to extrapolate how prevalent they are. However, from the Introduction we have seen that there are certainly still groups advocating to establish them, and we have seen that the initiative to start them can come from either or both parties. In this section I discuss what I view as the primary, higher-level findings of this work, followed by a section drawing conclusions as to what it means for government and areas for future research.

Open records policy, and, basically all information access policies, become resource allocation policies when they are used to obtain data with commercial value – and they are inadequate to this task. This point was made in the introduction to this study and recurs throughout. Essentially, once it has been determined that government information has a secondary economic value outside its acquisition for mission purposes, then any policy regarding it takes on the role of resource allocation. If all it says is "it's free, take it," then everything beyond that appears to be left, in most cases, to the discretion of administrative decision makers. Whether information must be requested or is put on websites, made available for download in raw form, or even designed to delivered real-time (such as satellites) is a broad palette of options from which to administer a government resource without further controls or guidance. Even when we see policies ostensibly designed to provide that guidance in the case on NOAA Environmental Data Policy, we find that it establishes guidelines that favor industry and creates channels for their input into decision making, leaving much of the mechanics of distribution negotiable. In essence, practice is policy where no policy exists, something we see in each case in this study.

The challenges presented by this lack of guidance are illustrated further below, but a critical flaw in the use of open records policy for this purpose is its inability to address the fact that *equal access* is not identical to *equal capabilities to access*, making distribution choices and technologies matter more as the size and velocity of data, and its consumption, grows. For static information, say, a government memo, both industry and an individual can download it via a personal computer and an internet connection. But, even among businesses, not all of them have the ability to tap into a high-speed real-time feed of radar or satellite data, or, if they can, store it or dynamically process it and deliver products. Or, perhaps in some cases, the government does not have sufficient technical capacity to serve unlimited users in the timeframes they need – one of the professed drivers of the NOAA Big Data Project concept – yet prioritization of users would lead to inequity. The lack of policies beyond just granting open access makes these decisions difficult, leaving the potential to introduce inequitable or unintended outcomes. The complexity of developing strategy and policy here is further compounded by the fact that private sector incentives for data access are not monolithic across an industry, nor are they driven by the benefit for all industry, but the benefit for the individual company. We see this in the Reston project at the end of the NOAA Environmental Data Policy chapter where vendors squabble about which will managed a shared line into and out of the high-speed, real-time data, where a few second priority may translate into money.

When government provides data to the private sector, it is participating in an information market, with all the attendant volatility and complexity. Dependence by government on the benefits from these relationships has real risks. This is reinforced in every case in this study and is significant for several reasons, most importantly when government comes to depend heavily on some benefit from this data provision to the private sector. This is because of the dynamic nature of the information market and – in all of the cases described – private use of the data is entirely optional, not a source of revenue based on compulsory participation like citizen payment of vehicle registration fees or income taxes. So, like any market, substitute goods, price volatility, the opacity of strategies and motivations of participants, technology innovations or policy and regulatory changes, the behavior and requirements of markets downstream – all these can change over time and serve to shape the viability of the data-providing relationship, its terms, and benefits to each party.

This risk is acknowledged explicitly, for example, in the interview with a manager who suggests the state won't take the risk of developing alternative products for sale to industry that could increase – but might endanger – their overall revenue from driver history record sales because the resulting funds are put to a highly visible use – law enforcement retirements. While they attempt to accommodate this risk in their strategy in this case, the sophistication of information markets and their dynamic nature means it is unlikely governments have the skills to manage their participation effectively. And, it is also apparent from that case that decision-making authority can be distributed in such a way that makes it difficult for them to respond quickly and in a coordinated fashion to market intelligence when they do receive it. Finally, the intent of the parties involved may be opaque. As we saw in the Big Data Project, at least one major vendor considered NOAA's goal to be as cover only to test the market to eventually purchase cloud services, not to really make a go of the proposed financial model. In turn, it was clear that NOAA did not have a good understanding of the motivations and organizational barriers inside the cloud providers implementing its model. And to this day, the motivations for those parties involved in the change of a state driver privacy protection statute remain unclear.

There are a myriad of barriers to successful policy feedback between government and the private sector in the secondary use of government information. The assumption was made early on, and still appears correct from this study that the primary benefit to industry from these relationships is revenue. It does appear, too, that for companies with multiple lines of business that could have sales opportunities in other areas, there may be motivation to participate just to maintain good relations with government and "stay in the game." However, for government, while revenue, when it can be had, looks to be a strong force for dependence and openness to influence, we see there can be other benefits including cost-avoidance (a form of revenue preservation) in performing their mission, as we see with some NOAA initiatives like NEXRAD and with genealogical records at the archive. That said, there are many barriers, some unexpected:

- **Dueling imaginaries.** It turns out that government may, in fact, not have a good understanding of information markets and private sector business models and influences. This is most evident in the NOAA Big Data Project. In turn, as we see with Restaurant Inspections, a company's assumption that government would get on board failed to include – even self-admittedly in the interview we listen to – a basic understanding of the motivations and technical capacities of government restaurant inspection units. The point here isn't the differences (that's my next point), but that there are certainly cases where both industry and government have plans for each other that appear to be based more on unverified assumptions and stereotypes than the actual fact son the ground.
- **Not made for each other.** This is a coy way of saying that, and here is the surprise, both government and industry, especially companies in those industries who have not previously used government data as part of their business model, are stymied in the secondary use of data by path dependence of organizational and process design, technology investments, and perhaps yet-to-be-discovered incompatibilities when expanding the study to a larger sample size. As we saw especially in the NOAA Big Data Project, private sector companies may not

be based on business models, be organized to, or recognize or respond to opportunities presented by government in this area. If they are organized only to sell products or services to government (the vast majority), this area may be confusing to them. This may be part of the rationale for the insertion of intermediaries. Or, in the case of cloud services, we see that some vendors have made their investment in technologies and that "all clouds are not created equal," so this makes both the calculus of switching or of using two technologies in parallel financially untenable. On the government side, we see that decision making in both restaurant inspections and in the sale of driver records may be distributed – there may be internal conflicts, and no unified place to exercise influence to resolve them. This also may be true of business, where, as we saw, it is not even clear to a company employee which business units are already using data. In these cases there are also anecdotes about government lacking the technology or expertise to actively deliver the goods, resulting in intermediaries stepping in to make the translation. This is another area for future research, especially given that these organizations are likely to remain opaque to each other, which brings me to the final aspect.

- **Lack of understanding and visibility of agency data and its private sector value by both parties**. Industry doesn't have a way to understand what data agencies have, and agencies may not have a catalog or inventory that is set up in a way – if it were possible – for either of them to determine which data sets are likely to have commercial value. The back-and-forth described by the participants in the NOAA Big Data project ("We thought you knew!") demonstrates this in a large agency with a vast quantity and types of data, but it seems likely to be the case in government agencies generally.

The takeaway here seems less how these projects, programs, and policies came out than what they say about the uncharted territory of the secondary market for government information and the missing policies, strategies and dialog on both sides that would be needed to align and exploit these relationships more fully for mutual benefit, while, on the government side, managing the democratic issues that come with dependence.

Mutual dependence on government information exists, it can affect government, and government can be the protagonist in influence, acting in anticipation of what it thinks the market might want. We see this primarily in the Big Data Project, but also in the NCEI data promotion program. As noted in the findings, there are impacts of many types on government in these relationships, whether it be working against legislation that would reduce their revenue, or setting pricing based on what the market will bear while tying critical programs to the revenue. Rights to public information are, at a minimum, traded away for services in the short-term, while, in the Big Data Project, they appear to trade expertise for free hosting. On the other side, pressure is brought by industry for a new government restaurant re-inspection service to be dropped to preserve the value of the data they profit from, a U.S. senator introduces legislation as a bargaining chip that would dramatically reduce the mission of the weather service, and industry agrees to participate in projects that they believe will go nowhere in the hopes of making money off something else, while government happily goes along, in part, legitimizing their innovative posture and publicizing the value of their programs.

Government dependence on private sector use of data shows signs of being a gateway for market values to seep into decisions made during the lifecycle of government information. While positioned as a final observation in this discussion, it is no less important than the policy vacuum around these relationships. We see the main signs of this in the NOAA data promotion case, where the concepts of “use-inspired” on the front line and “societal benefits" at the strategy and investment-setting level are primed to consider private sector requirements and benefits to the publics they serve in the process of making decisions about what scientific work to prioritize, or possibly undertake. In NOAA's case specifically, increasing financialization of natural resources and their data's potential value in future schemes of natural capital accounting look to entwine these measurements further in the market and its needs. But, this is not unique to this case. In driver history records, we see legislation being stymied (or supported) in the name of preserving or increasing government revenue, with public values seemingly of less (or no) concern in the discussion. This influence is hypothesized in the case of archives as well, where, in practice – even if they do retain rights for the citizens of the state (hard to believe everyone still lives in the state in which they were born) – they potentially trade away practical access to certain records by the public to the pricing decisions of commercial parties. While this concept is akin to concerns raised when government functions are outsourced to the private sector, the process here is more subtle, and the stakeholder relationships such that they may be optional for business, targeted only to areas of their interest, perhaps shaping what information is practically available to the public, but ones upon which government may depend for its mission. As a core concern of the problem statement behind this research, the signs are quite sobering.

Conclusion and Recommendations for Further Research

In the introduction to this study, I expressed a real concern that mutual dependence between industry and government might shape the very nature and composition of public data. I feared that, at the extreme, processes of mutual dependence may erode the public value of governmental data to serve private purposes. Although my research did not discover distortions at that extreme end of the spectrum of possibilities, the experience conducting this study has only reinforced my concern. In many of the cases studies here, processes of mutual dependence are abundantly evident, and subtle influences on the process of generating and distributing governmental data are widespread. One can take refuge, I suppose, in the numerous barriers to fundamental corruption of the data that I have identified. Still, these barriers mainly take the form of path dependency (that is, organizational resistance to change) rather than deliberately-crafted policies or procedures to protect the public interest. But my experiences in investigating this subject show that this potential impact remains, and it may be as simple as developing strategies in policy design to overcome or mitigate some of the barriers I identified for these impacts to grow, with governments perhaps willing to take more risks as revenue and other benefits become harder to come by. However, I don't think that's enough refuge, as the problem doesn't have to be universal in government to qualify as a problem, or, if unaddressed, a threat.

The amount of government data available for secondary use is growing. With advances in computer processing power, network bandwidth, and data collection (through sensors and satellites – or as NOAA says, "from the bottom of the ocean to the surface of the sun"), government is interested in and produces more and more data at every level. Data for use in decision-making, in monitoring, in performance measurement, in resource management and allocation. And yet we also see budget and resource challenges, both to support investments in this data collection and management, but also in the programs and operation of government to deliver this data on behalf of its citizens. In the big picture, we see very large commercial interests, global in scale – some of whom participated in this study or were identified in NOAA's "Success Stories on User Engagement (Success, 2018) – who have interests in data of many types. Even though use (versus hosting) of government data may not be attractive to them currently, many businesses who are their clients or partners are quite interested. And, as we've seen in the cases where money is involved, private interests will come to the table to defend them.

To address this policy vacuum requires further research, but will also operate with an overarching constraint. As was noted to me recently, these secondary uses of data are curiously non-ideological – and so they are. This seems to me to echo the classic symbolic legacy referred to by Beland (2010) that can facilitate or deter feedback effects. In this case, transparency, openness, and access to data the taxpayer has already paid for is so strong that any attempts to further regulate or shape private sector use of it seems destined to fail as an attempt to reduce accountability of government – even when, in this case, the intent would be to make government more accountable. In turn, this constraint in developing policy will inevitably bump up against another: The simple lack of resources, much less the mandate, for government to be able to make all this data available. As the figures calculated by one of the vendors quoted in the NOAA Big Data Project showed, just storing the data alone, not processing or using it, would become astronomical in cost in the near term (PRI-20). So, the situation, at least in the case of environmental information, seems untenable. Most likely, government and industry will stumble along, using paths they are already both familiar with to resist further policy and regulation, opportunistically seeking benefits where they are to be gained on either or both sides, and the problem will continue, unmonitored or cataloged. And, perhaps due to this, less researched or discussed.

It is with this assessment in mind that I propose directions for future research. At this point, other than raising awareness of these issues, I have not yet collected enough evidence to offer prescriptive policies or recommendations that would counteract the effects of dependence that we have seen in these cases. So, it seems that the road begins with continued research on questions posed by this analysis, along with some suggested by its limitations.

The key question in this study involves the dependence of government on benefits they derive from the private sector's use of their data. While in a case where we see large amounts of revenue received by government, perhaps it is easy enough to understand why they would quickly come to depend on or even seek to enhance it, and work against threats to its continuance. And, policy feedback theory has been useful in understanding how private sector users of this data coalesce around it and seek to gain further benefits themselves in the form of concessions from government that lower their costs to obtain it, increasing the reliability of its supply, and restraining government from competing with them in its use. On the other hand, why do governments seek out these relationships or benefits? And why might they be willing to adjust the focus of their operations, redirect resources, or even change the format or content of their data to obtain them?

Three theories have promise for addressing these questions, as well as shedding further light on the motivations and organizational aspects of mutual dependence. The first is the literature of Bureaucratic Politics, a set of theories that seeks to explain governmental decisions as the products of political bargaining and coalition-building among competing bureaucratic actors (O'Leary, 2020). The core of the theory, as formulated by Allison (1969; 1971) is the observation that key decisions at the highest levels of government are made collectively by bureaucratic actors and that these actors often differ in their goals and perspectives and must negotiate or bargain to reach a decision. Allison built on a long tradition of scholarship observing that political bargaining is as central to bureaucratic processes as to nearly any other governmental process (see, e.g., Neustadt 1960). The theory has since been expanded to address bureaucratic politics at lower levels of the bureaucracy, and the observation that bureaucratic actors differ not only in their goals but also their knowledge, the information available to them, and their power (Bendor and Hammond 1992; O'Leary 2020). The model's enduring contribution is to focus attention on how bureaucratic decisions are often negotiated or bargained via a process that is political in its basic elements.

The first recommendation for further research, therefore, is to draw on the insights of the bureaucratic politics model to further study how processes of bargaining and negotiation shape what I have called mutual dependency. This study is replete with examples of bargaining among governmental actors and industry officials, in the context of conflicting goals, differing knowledge, and differences in power. In the DHR case, executives recount negotiating with other agencies, or even other branches of government, to preserve their revenue, while legislators turn a blind eye to its intended use to fund other priorities. Health officials attempt to maintain good relations with restaurants whose fees pay their salaries and institute an instant re-inspection program to remain in their good graces in the face of a company's public shaming. The NCEI invests resources to document and promote the value of its data to industry to build support for its programs and their funding, while policy entrepreneurs in the Big Data Project develop plans to disrupt the market with what they believe to be an innovative model for business that can help salvage their own budget for data storage and delivery.

Where the theories of policy feedback and path dependency used to organize my study drew attention to the often unplanned and non-deliberate ways in which mutual dependency shaped the process, the theory of bureaucratic politics would turn attention specifically to the deliberate bargaining and power-politics that may also shape the process. While I hope to have lent support to the proposition that processes of mutual dependency in the area of governmental data are heavily path-dependent, it is also surely true that some aspects of the process are amenable to deliberate control and frank negotiation or conflict among actors with competing interests and differing knowledge. Thus, government executives are experienced professionals with deep commitments to their agencies' missions and, often, to the need for revenue or political support to carry on those missions. It is abundantly clear that many recognize that revenue and political support may be enhanced via making governmental data available to private industry. While they might not have a good understanding of the market (bounded rationality), they certainly have a sense of the tradeoffs, options, and policy agendas that surround their daily operations – and appear eager to trade upon the value (actual or symbolic) of the data they have, either inside or outside the organization, if the opportunity looks promising. Even NOAA, a science agency animated by expertise and accuracy, exhibited a willingness to expand the scope of "use-inspired" products to be more responsive to industry, expanding their focus or increasing their frequency of production. On the industry side, it is certainly true that industry officials are highly attuned to their corporations' goals of making money. So, the processes examined in this dissertation exhibit some aspects of the bargaining relationships that are the province of bureaucratic politics theory. Examining these and future cases through the lens of that theory may illuminate aspects of these processes not fully revealed by my focus on policy feedback and path dependency.

A second theory that may help to illuminate other aspects of the process is Resource Dependence Theory (RDT) (Pfeffer & Salancik, 2003). This theory recognizes that organizations depend on resources from outside their boundaries for success, and, because of this, become subject to both mutual dependence with suppliers of these resources, and also come to participate in a network of external interdependencies between them and other organizations.

While this theory developed primarily to explain relationships between private companies (Hillman, Withers, & Collins, 2009), the incentives and forces in play in dependence on information as a resource may well be analogous in their logics and effects in the relationships between the public and private sector.

In the cases examined here, the benefits to government conferred by private sector use of its data may be revenue - at the state or local level, it is prohibited federally. This is a critical resource for obvious reasons. But, agencies may also benefit from the prestige or legitimacy conferred on the programs that create the data that comes from private sector use – which may help build legislative or public support for continued or increased funding or even expansion of agency programs, or it may help government avoid costs as it is used to deliver services an agency might otherwise be called on to provide. While the relative size of these benefits varies, it is likely that their importance also varies and, in some cases, may be critical resources to the agency. So, it is from this calculus that one might apply RDT to identify the behaviors we have seen, or new ones that should be explored further. While this study does not seek to provide in-depth explanation of mutual dependence on the private sector, one can assume that, at least in some cases described here, government information is a critical resource to industry and thus their behaviors might also mimic the expectations set out by the theory.

In RDT, the motivations of a party in government's position would be to find ways to reduce the uncertainty around receiving these benefits, and to also reduce - or perhaps diversify, their dependence. So, for example, when NOAA looks to broaden the use of their data, it can be seen as fulfilling their mandate – certainly of government open data policies – but might also be seen as an attempt to broaden their support, decreasing their reliance on any one party as a constituent. The emphasis in this theory on organizational forces and inter-organizational relationships may provide some insight into constraints on the range of actions predicted through the theories of bureaucratic politics discussed earlier, as, in RDT, the impact of managers is “[F]requently constrained by situational contingencies and the individual’s effect is relatively small” (Pfeffer & Salancik, 2003).

Conversely, the difference in the organizational structure of government vs. the private sector may make some aspects of the theory difficult to translate – for example, actions taken by companies to absorb (merge) with their suppliers to reduce uncertainty are not feasible, although perhaps that inclination might result in more contracting. In turn, where the literature suggests influence and stabilization via boards of directors, in government, the private sector may – and does, in the example of NOAA Environmental Policy in this study – use the legislative process or other mechanisms for public engagement to shape these relationships more to their liking.

The third theoretical perspective that may be helpful in analyzing mutual dependence uses the conceptual frame of “Iron Triangles” (Heclo & King, 1978), linking congressional committees, interest groups, and government agencies in a system of interdependence that allows industry to gain concessions from government in return for providing political support to legislators. The third side of the triangle then traces the path of funding and other support provided by the legislature back to the agency. In the case in this study focused on NOAA’s policy on environmental data sharing, we do see what appears to be something of this sort: NOAA attempts to appease industry by ensuring their input and including processes for NOAA’s ongoing responsiveness to it as part of a revised policy. Industry was not satisfied with this concession however, and we see cases where industry communicates this to NOAA, but also where a congressman introduces legislation to remove almost all barriers to industry access to NOAA data and privatize most of their functions. The policy changes, the legislation is abandoned, and NOAA funding continues. NOAA is not unique among federal agencies in this regard, and for example, the motivations and political influence that may have been involved at the state level in the DHR case where legislation was changed to favor expanded industry access remains unclear. By further exploration, then, of interest groups formed by or aligned with private sector users of government data using these concepts, there may yet be more to reveal of the motivations and techniques of industry and government in cultivating, maintaining, or even expanding mutual dependence.

In addition to research deploying different theoretical lenses than those used in this study, there is a need for studying how processes of mutual dependency may vary among different types of governmental agencies and different kinds of business interests. As we saw in this study, some agencies may be more in the "data business" than others. Cases involving different organizational structures may be helpful to more clearly identify the impact of, for example, the decentralized nature of decision making in the case of restaurant inspections, with organizations where they may be more centralized, like archives. Another aspect that may be germane is whether or not the agency produces primarily regulatory, scientific, or administrative data – or yet some other type - and how those types may be treated differently, or be subject to different commercial interests and strategies for access and use. The frequency of data collection and its value over time can be other important dimensions, with one-time records like those of interest to genealogists involving a different set of decisions and influences than data streaming real-time from weather satellites. And, while the data discussed here has been primarily "structured" (with defined fields and values), unstructured data in the form of reports, or open-ended narratives to which machine learning may later be applied may also present different challenges and opportunities for influence. In all these cases as well, the subject of pricing for data should be considered. It is not clear, for example, what forces restrain the federal government, who were once able to charge for data, from doing so again as is done in Europe. And, on the state level, where fees can be levied for the provision of data, their implementation is inconsistent, with some types of data being charged for under statutory authority to do so, and others not. This facet, too, will no doubt produce new insights of interest.

While there is certainly more to investigate inside government operations, it will be equally necessary to further explore the commercial ecosystems that have interest in the data they produce. Cases where there is competition, like in meteorological data, may differ from where there are few companies interested (apparently like genealogical records) or that dominate a market. One significant piece missing from this research is also more in-depth examination of the role of intermediaries that sit between government and commercial users, whether they be private firms (like the DHR case) or even an extension of government or academia (as in the CICS-NC in NOAA's Big Data Project). Variation in the detailed mechanics of their role in facilitating the exchange of data may uncover new incentives or behaviors on both sides, and activities they may perform to compensate for the mismatch of organization and mission between the two parties could provide more clues as to needed policies.

My final recommendation is for research aimed at developing a public information policy for the current era. The gap in this area is large. Policies governing access to governmental data mainly take the form of guarantees of access: these are the celebrated Freedom of Information Act at the federal level and the various open records laws in the states. Freedom of Information remains an important value. But it addressed the key problem of an earlier era: the inability of members of the public to learn the then-secret policies and information held by the government. Although aspects of that problem surely remain, this study has revealed a fundamentally different problem characteristic of our age: the shaping of governmental data to serve the purposes of private industry.

To address this problem, the public needs a new form of information policy whose purpose is to ensure the *public* integrity of governmental data, which is to say, to ensure that governmental data's basic purpose is to serve the public interest. Current policy is inadequate to address this need. No overarching policy favors the public integrity of governmental data. Considerable research will be necessary to develop it.

That is in part because the types and character of governmental data are so varied, and are generated by widely varying processes in differing types of agency settings. More fundamentally, however, the challenge is deepened by the need to subtly distinguish between the legitimate interests of private businesses and government agencies in enhancing availability of data for business purposes that may benefit the public, on the one hand, from, on the other, corruption of these processes in ways that risk public value and data integrity. To return to the problem statement that gave rise to this study, recall that Lessig (2013) clearly defines the situation we wish such policy to address:

> Institutional corruption is manifest when there is a systemic and strategic influence which is legal, or even currently ethical, *that undermines the institution's effectiveness by diverting it from its purpose or weakening its ability to achieve its purpose*, *including, to the extent relevant to its purpose, weakening either the public's trust in that institution or the institution's inherent trustworthiness*. [emphasis added]

A possible way to focus policy reform in this area may be Overman and Cahill's (1990) suggestion that, among the many values shaping governmental data—e.g., costs, secrecy and the public's right to know—is the data's "*usefulness*." In their view (1990), [T]he key policy issue surrounding the usefulness value revolves around who will decide and control what information is to be considered useful and therefore collected and stored by the federal government."

Policies aimed at addressing the public integrity of governmental data should begin here, by enhancing public discussion and oversight of the data's "usefulness." These policies should aim to ensure that usefulness is defined by *public* values and not primarily by private interests. First steps in this direction may include policies that require agencies to:

- Publicly identify known uses of agency data for commercial purposes.
- Make visible specific benefits the agency receives through external use of its data as well as those upon which it relies for program success.
- Segregate and report revenue and cost-avoidance obtained through use of agency data
- Specifically identify and report categorically any contracts or agreements with external parties, primarily commercial firms, to provide data for what are apparently commercial uses, including those that involve the receipt of non-cash benefits such as services as compensation for this use.

While these elements are not mutually exclusive, they can assist in identifying areas where influence has a greater likelihood of occurring, as well as provide data for use in further academic research in this area that can further inform policy. The foregoing steps aim to enhance public transparency regarding what organized interests have a stake in the utility of governmental data.

Another category of policies might go further by enhancing public oversight over the process of defining this utility. One can imagine "public data integrity committees" whose mission is to exercise oversight over any resale or other reuse of governmental data by private businesses. These initial areas of reform only scratch the surface of what may be necessary. Research on these matters should build on the growing body of scholarship on how data are shaped in an ongoing process from conceptualization of the need for information through constructing the instruments to gather the data and on to archiving it in the end. This process might be called the "information stack" or "information lifecycle." As was intentionally demonstrated in the NEXRAD case, we know that opportunities for choice – and influence – may be found throughout the information lifecycle, from creation or collection of data, to retention, processing, and eventual access and distribution methods. Because of this, any approach to addressing the policy vacuum should focus on the full lifecycle and integrate the resulting policies to work together to ensure the public interest is served. As I noted at the start of this section, however, I believe more research is needed simply to better understand where public values and interests may apply in choices about information.

Growing scholarly interest in the construction of information, its power, and uses has coalesced in the new interdisciplinary field of Critical Data Studies (CDS). One unfilled promise in this study is a much closer look at the process of construction of government information with an eye toward understanding the opportunities for influence and the methods of its application specifically in that process. In CDS, I believe the questions posed in this research can find a home. A recent summary of the way this field looks at data fits closely with the assumptions underlying this research, especially those related to information construction:

> Data are a form of power. Organizations own vast quantities of user information and hold lucrative data capital (Yousif, 2015), wield algorithms and data processing tools with the ability to influence emotions and culture (Gillespie, 2014; Kramer et al., 2016; Striphas, 2015), and researchers invoke data in the name of scientific objectivity while often ignoring that data are never raw but always "cooked" (Gitelman, 2013). There is evidence that data are surreptitiously extracted from data subjects (Hauge et al., 2016; Metcalf and Crawford, 2016), hijacked to serve agendas that benefit research and industry (Ioannidis, 2005, 2016), and compromised by the interests of not only powerful business organizations but also hackers and rogue agents (Coleman, 2014; Elmer et al., 2015). While data are all of the above and more, they are also conspicuous in their absence—a lack of data is another indication of power, the power not to look or to remain hidden (Brunton and Nissenbaum, 2015; Flyverbom et al., 2016). In their presence and absence, data are always-already active and never neutral, part of an information geography (Graham, 2014, 2015) that is always in flux. (Iliadis & Russo, 2016)

To date, this research agenda already incorporates work on government information, including the work of Bates (2014a, 2014b) on commercialization of climate data and neoliberal impacts on information policy in the U.K. There is a natural bridge to research focused on bringing public accountability via standards for documenting how data is collected and underlying facets of its structure and meaning in anticipation of its use by algorithms and artificial intelligence (See "Datasheets for Datasets", Gebru et.al, 2018). That this is an important area for government is signaled by recent work showing bias in algorithmic systems and data-based decision making in the public sector (Eubanks, 2018).

In a timely development that parallels this new field of research, a new investment is being made by the federal government that focuses on addressing data as a resource. The Federal Data Strategy: Leveraging Data as a Strategic Asset (OMB, 2020a), is a cross-agency program with a mission to "[F]ully leverage the value of federal data for mission, service, and the public good by guiding the Federal Government in practicing ethical governance, conscious design, and a learning culture." This effort is already well underway, supported by a consensus-based set of 40 practices for its execution (OMB, 2020b). Yet again, as with the existing open data policy (EOP, 2013), the focus remains on support for commercial dependence. Practice 10, for example, includes the call to "Ensure that sufficient human and financial resources are available to support data driven agency decision-making, accountability and the ability to spur commercialization, innovation, and public use" (OMB, 2020b).

With these parallel efforts in government and academia, it appears less feasible to propose a wholesale replacement of government information policy than to attempt to marry the two interests together, leveraging more academic interest in the subject of influence throughout the information lifecycle with an eye toward integrating the results of research into the framework of guidelines already being developed by government. In turn, as these policies mature, similar approaches may be translated to state and local government, where the sale of data is involved, and adjusted as needed to accommodate that variation in dependence.

* * *

My suggestions for further research illustrate how much there is yet to learn about private reuse of governmental data and its implications for government and the public. As an information technology professional with a career of public service, that acknowledgment is more than a bit humbling. Although I believe that I know this area of policy and practice well, my studies have revealed how complex these interactions can be, and how little is known in any systematic way about them and their effects.

In my explorations of this subject, I have had the opportunity to see a rich world of interaction between government and the private sector that may be invisible to many in the world outside it. My hope is that this study represents a first step in a research agenda that further engages the areas outlined above and I look forward to using the results to develop both principles and guidance that will be useful to public administrators in identifying and successfully negotiating the challenges of information dependence.

References

151 Cong. Rec. (Bound) 6568 (2005) (statement of Sen. Rick Santorum). Retrieved from https://www.gpo.gov/fdsys/pkg/CRECB-2005-pt5/pdf/CRECB-2005-pt5-Pg6566.pdf

About the National Centers for Environmental Information (2018). Retrieved from https://www.ncei.noaa.gov/about-national-centers-environmental-information

About Us (2018). Retrieved from https://www.ncdc.noaa.gov/about.

Adopting the Driver's Privacy and Protection Act: *Hearing of the House Transportation Committee*, February 13, 2019. Retrieved from: http://sg001-harmony.sliq.net/00287/Harmony/en/PowerBrowser/PowerBrowserV2/20190313/1/6494

Advisory Committee on Weather Services (1953) Weather is the Nation's Business, Department of Commerce, U.S. Government Printing Office. Washington, D.C.

Alaska Administrative Code (A.A.C), 2020. Section 28.10.505 Disclosure of Personal Information Contained in Motor Vehicle Records. Retrieved from: http://www.legis.state.ak.us/basis/folioproxy.asp?url=http://www.legis.state.ak.us/cgi-bin/folioisa.dll/stattx08/query=*/doc/%7Bt12635%7D?

Allison, Graham. 1969. "Conceptual Models and the Cuban Missile Crisis." *American Political Science Review* 63:689-718.

---------- 1971. *Essence of Decision: Explaining the Cuban Missile Crisis*. Boston: Little, Brown.

American Institute of Physics. (April 6, 2018). Final FY18 Appropriations: National Oceanic and Atmospheric Administration. Retrieved from: https://www.aip.org/fyi/2018/final-fy18-appropriations-national-oceanic-and-atmospheric-administration

American Meteorological Society (1981) News from our chapters. *Bulletin of the American Meteorological Society*, 282-287, Volume 62, Number 2.

American Meteorological Society (1999). 15th International Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology Dallas, Texas. January 10-15, 1999. Retrieved from: https://ams.confex.com/ams/older/99annual/iips.htm

American Meteorological Society (2001). 17th International Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology. Albuquerque, New Mexico. January 13-18, 2001. Retrieved from: https://ams.confex.com/ams/annual2001/webprogram/17IIPS.html

American Meteorological Society (2004). 20th International Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology. Seattle, Washington. January 10-16, 2004. Retrieved from: https://ams.confex.com/ams/84Annual/techprogram/programexpanded_182.htm

American Meteorological Society (2010). 26th Conference on Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology. Atlanta, Georgia. January 16-21, 2010. Retrieved from: https://ams.confex.com/ams/90annual/techprogram/programexpanded_583.htm

American Meteorological Society (2012). State of the Weather and Climate Enterprise. Retrieved from: https://www.ametsoc.org/index.cfm/cwwce/reports-and-studies/general-reports-and-studies/state-of-the-weather-and-climate-enterprise-in-2012/

American Meteorological Society Commission on Weather, Water, and Climate Enterprise (2014b). Committee Minutes of the AMS CWCE BEED Committee on Open Environmental Information Services (COEIS) - Monday, February 3, 2014. Retrieved from: https://www.ametsoc.org/index.cfm/cwwce/committees/committee-on-open-environmental-information-services/meeting-minutes/minutes-of-the-coeis-feb-3-2014-meeting-at-the-ams-annual-meeting/

American Meteorological Society Commission on Weather, Water, and Climate Enterprise (2014). Report to Council – December 2014. Retrieved from https://www.ametsoc.org/index.cfm/cwwce/committees/cwwce-executive-committee/reports-to-council/cwwce-report-to-council-december-2014/

American Meteorological Society Commission on Weather, Water, and Climate Enterprise (2014). Report to Council – December 2014. Retrieved from https://www.ametsoc.org/index.cfm/cwwce/committees/cwwce-executive-committee/reports-to-council/cwwce-report-to-council-december-2014/

Ansari, S., Del Greco, S., Kearns, E., Brown, O., Wilkins, S., Ramamurthy, M., Weber, J., May, R., Sundwall, J., Layton, J., Gold, A., Pasch A., and Lakshmanan, V. (2018). Unlocking the potential of NEXRAD data through NOAA's Big Data Partnership. *Bulletin of the American Meteorological Society*, 99(1), 189-204.

Arguez, A., Vose, R., Dissen, J. (2013). Alternative Climate Normals: Impacts to the Energy Industry. Bulletin of the American Meteorological Society, June 2013. Retrieved from https://journals.ametsoc.org/doi/full/10.1175/BAMS-D-12-00155.1

Arkansas Code Annotated (A.C.A.), 2020. § 27-50-906 Furnishing of abstracts — Definition. Retrieved from: https://advance.lexis.com/documentpage/?pdmfid=1000516&crid=b6c9b73c-7d94-49a5-8658-e27ec4995ebc&nodeid=ABBAAEAACAAJAAH&nodepath=%2FROOT%2FABB%2FABBAAE%2FABBAAEAAC%2FABBAAEAACAAJ%2FABBAAEAACAAJAAH&le

vel=5&haschildren=&populated=false&title=27-50-906.+Furnishing+of+abstracts+%E2%80%94+Definition.&config=00JAA2ZjZiM2VhNS0wNTVlLTQ3NzUtYjQzYy0yYWZmODJiODRmMDYKAFBvZENhdGFsb2fXiYCnsel0plIgqpYkw9PK&pddocfullpath=%2Fshared%2Fdocument%2Fstatutes-legislation%2Furn%3AcontentItem%3A5GJ2-NY10-R03K-929K-00008-00&ecomp=h3t7kkk&prid=a63d9e6d-1f89-4a77-886f-59c7c4e0fc34

Arnold, R. (1992). *The Logic of Congressional Action*. New Haven, CT: Yale University Press.

Arthur, W. (1989). Competing technologies, increasing returns, and lock-in by historical events. *The Economic Journal*, 116-131.

Associated Press. (2014 September 5). Court decides Yelp can change ratings. *New York Times*. Retrieved from: https://www.nytimes.com/2014/09/05/business/court-decides-yelp-can-change-ratings.html

Association of Food and Drug Officials (AFDO) (2017 October 16). Partners with a Common Purpose –A New AFDO and Industry Vision [Press Release]. Retrieved from: http://www.afdo.org/resources/Documents/4-news-and-events/2017-10-16-Partners_With_Common_Purpose.pdf

Baer, V. E. (1991). The transition from the present radar dissemination system to the NEXRAD Information Dissemination Service (NIDS). *Bulletin of the American Meteorological Society*, 72(1), 29-33.

Bates, J. (2014a). The strategic importance of information policy for the contemporary neoliberal state: the case of Open Government Data in the United Kingdom. Government Information Quarterly, 31(3), 388-395.

Bates, J. (2014b). SPERI Paper No. 13: Climate Risk, Big Data, and the Weather Market. Sheffield Political Economy Research Institute. Retrieved from http://speri.dept.shef.ac.uk/wp-content/uploads/2014/05/SPERI-Paper-No.-13.pdf

Bauer, B. (2015 August 13). Your Family's Genealogical Records May Have Been Digitized by a Prisoner. *Mother Jones*. Retrieved from: https://www.motherjones.com/politics/2015/08/mormon-church-prison-geneology-family-search/

Béland, D. (2010). "Reconsidering policy feedback: How policies affect politics." *Administration & Society*.

Bendor, Jonathan, and Thomas H. Hammond. 1992. *American Political Science Review*, 86(2): 301-322.

Bigler, S. G. (1969). Dial a radar. *Bulletin of the American Meteorological Society,* 50(6), 428-430.

Bigler, S. G. (1981). Radar: A short history. *Weatherwise*, 34(4), 158-163.

Blacher, M. (2016). *PennDOT Selling Drivers' Personal Information? NBC10 Investigators*. NBC 10 Television. Philadephia, PA. Retrieved from: http://www.nbcphiladelphia.com/news/local/PennDOT-Selling-Drivers-Personal-Information--NBC10-Investogators-364903281.html

Blacher, M. (2016). *PennDOT Selling Drivers' Personal Information? NBC10 Investigators*. NBC 10 Television. Philadephia, PA. Retrieved from: http://www.nbcphiladelphia.com/news/local/PennDOT-Selling-Drivers-Personal-Information--NBC10-Investogators-364903281.html

Blackman, J., Maidenberg, M., Varnham, S. (2018, April 8). Mexico's disaster bonds were meant to provide quick cash after hurricanes and earthquakes. But it often hasn't worked out that way. Retrieved from http://www.latimes.com/world/mexico-americas/la-na-mexico-catastrophe-bonds-20180405-htmlstory.html

Borowitz, M. (2019). Government data, commercial cloud: Will public access suffer?. Science, 363(6427), 588-589.

Bower, J., Christensen. C. (1995) "Disruptive Technologies: Catching the Wave." Harvard Business Review 73, no. 1 (January–February 1995): 43–53.

Bowker, G. & Star, S. (2000). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press.

Bowker, G. (1996). The history of information infrastructures: The case of the international classification of diseases. *Information Processing & Management*, *32*(1), 49-61.

Bowker, G. (2005). *Memory practices in the sciences* (pp. 269-298). Cambridge, MA: MIT Press.

Bradsher, J. G. (1989). An Introduction to Archives. In. J.G. Bradsher (Ed.), Managing Archives and Archival Institutions (pp. 1-17). Chicago, IL: University of Chicago Press.

Brandeis, L. (1913). What Publicity Can Do. *Harper's Weekly*, December 20, 1913, pp. 10-13.

BRB Publications, Inc. (2016). *State Fees for Driving Records*. Retrieved from https://www.mvrdecoder.com/content/drvlicfees.aspx

Brewer, M. (2015) NCEI User Engagement Strategy [PowerPoint slides]. Retrieved from ftp://ftp.ncdc.noaa.gov/pub/data/sds/cdr/conferences/2015%20PI%20Annual%20Meeting%20-%20Presentations/Day_3/(G-1)%20NCEI%20User%20Engagement%20Strategy%20-%20(Brewer).pdf

Campbell, A. (2003). *How Policies Make Citizens: Senior Political Activism and the American Welfare State*. Princeton, NJ: Princeton University Press.

Campbell, A. (2012) "Policy Makes Mass Politics." *Annual Review of Political Science* 15, 333-351.

Carfax (2020) Carfax Vehicle History Reports. Retrieved from https://www.carfax.com/vehicle-history-reports/

Casey, C., Kearns, E., de la Beaujardiere, J. & Kihn, E. (2018) Big Data and Cloud Initiatives at NOAA. Committee on Earth Observation Satellites, Future Data Architectures Big Data Workshop, Boulder, Colorado, April 26, 2018 [presentation]. Retrieved from: http://ceos.org/document_management/Meetings/Future%20Data%20Architectures%20Big%20Data%20Workshop%20April%202018/NOAA%20Big%20Data%20and%20Cloud%20Initiatives.pptx

Catlaw, T. & Sandberg, B. (2014). “Dangerous Government” Info-Liberalism, Active Citizenship, and the Open Government Directive. *Administration & Society*, *46*(3), 223-254.

Center for Open Data Enterprise (2019). What we do. Retrieved from: http://opendataenterprise.org/what-we-do

Centers for Disease Control and Prevention (2016). Multistate Outbreaks of Shiga toxin-producing Escherichia coli O26 Infections Linked to Chipotle Mexican Grill Restaurants (Final Update), February 1, 2016. Retrieved from: https://www.cdc.gov/ecoli/2015/o26-11-15/index.html

Centers for Disease Control and Prevention (2016b) Highlights from the 2016 Surveillance Report. Retrieved from: https://www.cdc.gov/fdoss/annual-reports/2016-report-highlights.html

Certification of Data (2018). Retrieved from https://www.ncdc.noaa.gov/customer-support/certification-data

CICS 2012 Annual Report. Retrieved from: https://ncics.org/ncics/pdfs/annual/cics_annual_2012.pdf

CICS 2017 Annual Report. Retrieved from: https://ncics.org/ncics/pdfs/annual/cics_annual_2017.pdf

Citron, D. & Pasquale, F. (2014). The scored society: due process for automated predictions. *Washington Law Review, 89.*

Clark, T. (2016). *NEXRAD Strategic Plan*. NOAA Radar Operations Center. Retrieved from: https://www.roc.noaa.gov/WSR88D/PublicDocs/NEXRAD_Strategic_Plan_2016.pdf

Climate Corporation (n.d.). Innovating at the Intersection of Agriculture and Technology. Retrieved from: https://climate.com/about

Cochran, N., Gordon, A., & Krause, S. (1980) Proactive Records: Reflections on the Village Watchman. Knowledge: Creation, Diffusion, *Utilization 2*, September, 5-18.

Copernicus in Brief (n.d.). Retrieved from: https://www.copernicus.eu/en/about-copernicus/copernicus-brief

Council of State Archivists (2017). The State of State Records: A Statistical Report on State Archives and Records Management Programs in the United States. Retrieved from: https://www.statearchivists.org/download_file/view/831/2914/

Criste, L. (2018). Cloud Services Market Expected to Hit All-Time High. Bloomberg.com. Retrieved from: https://about.bgov.com/blog/cloud-services-market-all-time-high/

Crum, T. D., & Alberty, R. L. (1993). The WSR-88D and the WSR-88D Operational Support Facility. Bulletin of the American Meteorological Society, 74(9), 1669-1688.

Crum, T. D., Alberty, R. L., & Burgess, D. W. (1993a). Recording, archiving, and using WSR-88D data. Bulletin of the American Meteorological Society, 74(4), 645-654.

Data Access - NCEI (2018). Retrieved from https://www.ncdc.noaa.gov/data-access

Davies, T., & Bawa, Z. (2012). The Promises and Perils of Open Government Data (OGD). *The Journal of Community Informatics*, *8*(2). Retrieved from http://ci-journal.net/index.php/ciej/article/view/929/955

de Oliveira, M.B. (2013). On the Commodification of Science: The Programmatic Dimension. *Science & Education*, 22, pp 2463–2483.

Del Greco, S. (2003) *P5F. 1 A History for WSR-88D Level II Data Services at the National Climatic Data Center*. 31st International Conference on Radar Meteorology. Seattle, Washington. August 11, 2003. Retrieved from: https://ams.confex.com/ams/pdfpapers/63055.pdf

Deloitte Analytics (2013). *Open growth. Stimulating demand for open data in the UK*. Retrieved from http://www2.deloitte.com/content/dam/Deloitte/uk/Documents/deloitte-analytics/open-growth.pdf

Denis, J., & Goëta, S. (2017). Rawification and the careful generation of open government data. Social studies of science, 47(5), 604-629.

Department of Commerce (1985). *Environmental Information Summaries (C-10): Radar Weather Observations*. Retrieved from: https://repository.library.noaa.gov/view/noaa/14288/noaa_14288_DS1.pdf?

Department of Commerce, National Oceanic and Atmospheric Administration (2011). Federal Meteorological Handbook No. 11: Doppler Radar Meteorological Handbook No. 11 – Doppler Radar Meteorological Observations Part A – System Concepts, Responsibilities, and Procedures CM-H11A. Retrieved from: https://www.roc.noaa.gov/wsr88d/PublicDocs/PartA_FMH11_Rev11.pdf

Department of Commerce, Office of Inspector General (1991). Final Inspection Report on NEXRAD Software Risks (TA-100). U.S. Congress, House Committee on Appropriations. Subcommittee on the Departments of Commerce, J. (1992). Departments of Commerce, Justice, and State, the judiciary, and related agencies appropriations for 1993: hearings before a subcommittee of the Committee on Appropriations, House of Representatives, One Hundred Second Congress, second session. Washington: U.S. G.P.O. Retrieved from: https://babel.hathitrust.org/cgi/pt?id=uc1.aa0005242979&view=1up&seq=128

Department of Commerce, Office of Inspector General (1993). National Oceanic and Atmospheric Administration: NEXRAD Software Maintenance. Final Inspection Report No. SED-4559-3-001. Retrieved from: Retrieved from: https://www.oig.doc.gov/OIGPublications/NOAA-SED-4559-3-001.pdf

Dissen, J. & Houston, T. (2014) Building Engagement Efforts for Advancing Climate Data to Decisions. Presented at Carolinas Climate Resilience Conference, April 29, 2014. Retrieved from: https://pdfs.semanticscholar.org/presentation/abc3/256691413c30c9ba215797d622739b595773.pdf

Dissen, J., Brewer, M. (June 6, 2018) Webinar: The Application of NOAA National Centers of Environmental Information Data and Information to the Legal Community**.** American Bar Association Continuing Legal Education. No URL.

Driver and Vehicle Record Access Information (2020). Retrieved from: https://egov.maryland.gov/mva/

Driver's Privacy Protection Act (1994). 18 U.S. Code § 2721. Public Law No. 103-322, codified as amended by Public Law No. 106-69.

Droegemeier, et. al. (2000). Project CRAFT: A Test Bed for Demonstrating the Real Time Acquisition and Archival of WSR-88D Base (Level II) Data. Manuscript submitted for publication to the Bulletin of the American Meteorological Society, October 1, 2000. Retrieved from: http://kkd.ou.edu/BAMS%20Article%20DRAFT%20for%20Distribution.pdf

Edwards, P. (2010). *A Vast Machine: Computer models, climate data, and the politics of global warming*. Cambridge, MA: MIT Press.

Electronic Privacy Information Center (EPIC) (2019). *The Drivers Privacy Protection Act (DPPA) and the Privacy of Your State Motor Vehicle Record*. Retrieved from: https://epic.org/privacy/drivers/

Environmental Data Management Committee (2013) *NOAA Environmental Data Management Framework v1.0*. Retrieved from: https://nosc.noaa.gov/EDMC/documents/NOAA_EDM_Framework_v1.0.pdf

Eriksson, F. & Ostberg, K. (2009). The problematic freedom of information principle. The Swedish experience. *Freedom of Information. Open Access, Empty Archives*. London: Routledge.

Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Press.

European Commission (2013). Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the Re-use of Public Sector Information. Retrieved from: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:EN:PDF

Executive Office of the President (2013a), "Making Open and Machine Readable the New Format for Government Information," Executive Order 13642, May 9, 2013. Retrieved from: http://www.ofr.gov/OFRUpload/OFRData/2013-11533_PI.pdf

Executive Office of the President, (2014a). *Presidential Innovation Fellows: By the People, For the People*. Retrieved from: https://web.archive.org/web/20140902003050/http://www.whitehouse.gov/innovationfellows

Executive Office of the President, (2014b). *Presidential Innovation Fellows: Projects - The Third Round*. Retrieved from: https://web.archive.org/web/20140902003411/http://www.whitehouse.gov/innovationfellows/projects#section-round-3

Executive Office of the President, "*Making Open and Machine Readable the New Format for Government Information*," Executive Order 13642, May 9, 2013. Retrieved from: http://www.ofr.gov/OFRUpload/OFRData/2013-11533_PI.pdf

Executive Office of the President, Office of Management and Budget (2013b), Memorandum for the Heads of Executive Departments and Agencies: Open Data Policy, Washington, DC, May 9, 2013. Retrieved from: http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf

Executive Office of the President. (2009). "Transparency and Open Government," 74 *Federal Register* 4685, January 26, 2009. Retrieved from: http://www.gpo.gov/fdsys/pkg/FR-2009-01-26/pdf/E9-1777.pdf.

Executive Office of the President. (2009). Office of Management and Budget, *Memorandum for the Heads of Executive Departments and Agencies: Open Government Directive*, Washington, DC, December 8, 2009. Retrieved from: http://www.whitehouse.gov/omb/assets/memoranda_2010/m10-06.pdf

Executive Office of the President. (2013). "*Making Open and Machine Readable the New Format for Government Information*," Executive Order 13642, May 9, 2013. Retrieved from: http://www.ofr.gov/OFRUpload/OFRData/2013-11533_PI.pdf.

Executive Office of the President. (2013). Office of Management and Budget, *Memorandum for the Heads of Executive Departments and Agencies: Open Data Policy*, Washington, DC, May 9, 2013. Retrieved from: http://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf

Executive Order 13653 (3 CFR 13653) *Preparing the United State for the Impacts of Climate Change*. Retrieved from: https://www.federalregister.gov/documents/2013/11/06/2013-26785/preparing-the-united-states-for-the-impacts-of-climate-change

Federal Records Management Council (2019). FRMC Digitization Services Sample Pricing Table (05/19). Retrieved from: https://www.archives.gov/files/digitization-services-pricing-tables.xlsx.

Feldman, M. & March, J. (1981). Information in organizations as signal and symbol. *Administrative Science Quarterly*, 26(2), 171-186.

Freedom of Information Act of 1966, 5 U.S.C. § 552 (2012)

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumeé III, H., & Crawford, K. (2018). Datasheets for datasets. *arXiv preprint arXiv:1803.09010*.

Geoeconomy (June 23, 2018). Data distribution drive launches Copernicus into next decade of growth. Retrieved from: https://geoconomy.com/index.php/2018/06/25/data-distribution-drive-launches-copernicus-into-next-decade-of-growth/

Glackin, M. (2010) NOAA Climate Service: Science Serving Society. Retrieved from ftp://ftp.oar.noaa.gov/SAB/sab/sab/members/2010/july/pdf/SABMarch232010_MG_Submitted_updated.pdf

Global Science & Technology, Inc. (2016). Case Study 4: The Reinsurance Industry. Retrieved from: https://www.ncei.noaa.gov/sites/default/files/ncei-reinsurance-report-2016.pdf

Global Science & Technology, Inc. (2016b) NCEI Climate Products and Services Market Analysis: Power Sector Engagement. Retrieved from https://ncics.org/ncics/pdfs/events/energy-sector-resilience/Acclimatise_NCEI_Jan8.pdf

Global Science & Technology, Inc. (2017). Success Story 6: The U.S. Drought Monitor. Retrieved from: https://www.ncei.noaa.gov/sites/default/files/drought_report_final_nov30_2017.pdf

Global Science & Technology, Inc. (2017b). Case Study 5: Retail and Manufacturing. Retrieved from https://www.ncei.noaa.gov/sites/default/files/success-stories-on-user-engagement-case-study-5-retail-and-manufacturing.pdf

Global Science & Technology, Inc. (2018). Success Story 8: Weather Service Providers. Retrieved from https://www.ncei.noaa.gov/sites/default/files/weather_service_providers_final_508c.pdf

Global Science & Technology, Inc. (2018b). Success Story 7: Logistics and Transport. Retrieved from: https://www.ncei.noaa.gov/sites/default/files/transport_and_logistics_report_final_508_compliant_4.pdf

Google (2018). Google Earth Engine Explorer – NOAA Datasets. Retrieved February 10, 2019 from https://explorer.earthengine.google.com/#search/noaa

Government Accountability Office (1990). Environmental Data: Major Effort Is Needed to Improve NOAA's Data Management and Archiving (GAO Publication IMTEC-91-11). Washington, D.C.: U.S. Government Printing Office. Retrieved from: https://www.gao.gov/assets/150/149940.pdf

Government Accountability Office. (1991). Weather Forecasting: Cost Growth and Delays in Billion-Dollar Weather Service Modernization. (GAO Publication IMTEC-92-12FS). Washington, D.C.: U.S. Government Printing Office. Retrieved from: https://www.gao.gov/assets/90/89205.pdf

Gratz, J. B. (2005). Unlocking the treasure chest of level-II radar data: lessons in technology transfer policy for the atmospheric sciences. In First Symposium on Policy Research. Retrieved online from: https://www.unidata.ucar.edu/committees/stratcom/2006april/104479-1.pdf

Grossman, R. (2015) From Clouds to Commons. Retrieved from: https://web.archive.org/web/20160607140558/http://occ-data.org/OCC-Commons/

Group on Earth Observations (2018). Earth Observations for Ecosystem Accounting (EO4EA). Retrieved from https://www.earthobservations.org/activity.php?id=111

Guillen, J. (2010, July 11). Ohio collects millions selling driving records with your personal information. *The Plain Dealer*. Retrieved from http://www.cleveland.com/open/index.ssf/2010/07/ohio_collects_millions_selling.html

Guillen, J. (2010, July 11). Ohio collects millions selling driving records with your personal information. *The Plain Dealer*. Retrieved from http://www.cleveland.com/open/index.ssf/2010/07/ohio_collects_millions_selling.html

Gurstein, M. (2011). Open data: Empowering the empowered or effective data use for everyone? *First Monday*, *16*(2). Retrieved from: http://journals.uic.edu/ojs/index.php/fm/article/view/3316/2764

H. 2126, 2019 Gen. Sess. (Kan. 2019). HB 2126 Adopting the Driver's Privacy Protection Act (Bill History). Retrieved from: http://www.kslegislature.org/li/b2019_20/measures/hb2126/

H. 2179a, 2019 Gen. Sess. (Kan. 2019). Retrieved from: http://www.kslegislature.org/li/b2019_20/measures/documents/hb2179_00_0000.pdf

H. 2179b, 2019 Gen. Sess. (Kan. 2019). HB 2179 Adopting the Driver's Privacy Protection Act (Bill History). Retrieved from: http://www.kslegislature.org/li/b2019_20/measures/hb2179/

Hacker, J. (2002). *The Divided Welfare State: The battle over public and private social benefits in the United States*. Cambridge, UK: Cambridge University Press.

Haselden, Leanne (2016). *NOAA Big Data CRADA – An IBM Perspective* [presentation]. From Joint Panel Discussion 1:The NOAA Big Data-CRADAs. An Update and a Look at the Opportunities They Pose for the AMS Community and Its Members, January 12, 2016 96th American Meteorological Society Annual Meeting, New Orleans, Louisiana, January 12, 2016. Retrieved from: https://ams.confex.com/ams/96Annual/videogateway.cgi/id/32872?recordingid=32872&uniqueid=Paper295161&entry_password=514389

Heclo, H., & King, A. (1978). Issue networks and the executive establishment. *Public Adm. Concepts Cases*, *413*(413), 46-57.

Hedstrom, M. (2002). Archives, memory, and interfaces with the past. *Archival Science*, *2*(1-2), 21-43.

Heimstädt, M. (2017). Openwashing: A decoupling perspective on organizational transparency. Technological Forecasting and Social Change, 125, 77-86.

Hillman, A. J., Withers, M. C., & Collins, B. J. (2009). Resource dependence theory: A review. *Journal of Management*, *35*(6), 1404-1427.

Hilton, R. P., & Hoag, N. E. (1966). Weather radar remoting system WB/RATTS-65 radar to telephone transmission system. In Proc. 12th Conf. on Weather Radar (pp. 91-97).

HIS Markit, 2019. Testimony in Support of HB 2179. House Transportation Committee, February 13, 2019. Retrieved from: http://www.kslegislature.org/li/b2019_20/committees/ctte_h_trnsprt_1/documents/testimony/20190213_01.pdf

IBM (2018). IBM and the Big Data Project. Retrieved from https://noaa-crada.mybluemix.net/node/3

IHS Markit, 2020. Carfax is part of IHS Markit. Retrieved from: https://ihsmarkit.com/btp/carfax.html

Iliadis, A., & Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society*, *3*(2).

Jain, M., & Rhue, D. (1995). WSR-88D's live data stream now accessible in real-time. *NSSL Briefings*.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, *29*(4), 258-268.

Johnson, E. (2001). Weather Radar Dissemination: A Policy and Technology Case Study. National Weather Service Presentation.

Johnson, E. (n.d.) Lessons from the National Weather Service Real-Time Level 2 Radar Data Distribution System. Internal document.

Johnson, J. A. (2014). From open data to information justice. *Ethics and Information Technology*, *16*(4), 263-274.

Jones, M. L. (2018). Ctrl+ Z: The right to be forgotten. NYU Press

Jorgensen, R. C., & Gerdes, W. F. (1951). The Dow Chemical Company leads industry in using radar for hurricane detection. Bulletin of the American Meteorological Society, 32(6), 221-224.

Kamerer, D. (2014). Understanding the Yelp review filter: An exploratory study. *First Monday*. Retrieved from: https://firstmonday.org/article/view/5436/4111

Kansas Historical Society (n.d.). Scheduling Electronic Records. Retrieved from: https://www.kshs.org/index.php?url=p/scheduling-electronic-records/11335

Kansas Statutes Annotated (K.S.A.), 74-2012. Retrieved from: https://www.ksrevisor.org/statutes/chapters/ch74/074_020_0012.html

Kearns, E. (2016). *Challenges and Opportunities with NOAA's Big Partnership Data.* Southeast Coastal Ocean Observing Regional Association (SECOORA) Stakeholder Meeting, Raleigh, North Carolina, May 19, 2016 [presentation]. Retrieved from: https://web.archive.org/web/20160712001413/http://secoora.org:80/webfm_send/1760

Kearns, E. (2017) Improving Access to Open Data through NOAA's Big Data Project. NIST Big Data Workshop Gaithersburg, MD, Friday June 2, 2017 [presentation]. Retrieved from: https://bigdatawg.nist.gov/Day2_08_NIST_Big_Data-Kearns.pdf

Kearns, E. (2018). *NOAA's Big Data Project*. U.S. Committee on the Marine Transportation System (CMTS) [webinar], April 20, 2018. Retrieved from: https://www.cmts.gov/downloads/CMTS_Webinar_Speaker_Kearns_20Apr2018.pdf

Kearns, E. (2018a) NOAA Big Data Project. NOAA Datafest, August 7, 2018 [online presentation]. Retrieved from: https://www.noaacrest.org/news-and-events/detail/2018-noaa-datafest

Kearns, E. (2018b) NOAA's Big Data Project. U.S. Committee on the Marine success System (CMTS) Webinar, April 20, 2018 [online presentation]. Retrieved from: https://www.cmts.gov/downloads/CMTS_Webinar_Speaker_Kearns_20Apr2018.pdf

Kelleher, K., et. al (2007). Project Craft: A real-time delivery system for NEXRAD Level-II Data via the Internet. Bulletin of the American Meteorological Society, 88(7), 1045-1058. Retrieved from: https://journals.ametsoc.org/doi/pdf/10.1175/BAMS-88-7-1045

Knapp, B & Corrin, A. (2018) GAO shuts down Oracle's JEDI protest, but it's not over yet. *Federal Times*. Retrieved from: https://www.federaltimes.com/acquisition/2018/11/15/gao-shuts-down-oracles-jedi-protest-but-its-not-over-yet/

Koblinsky, C., Pulwarty, R., Davidson, M. (2009) NOAA's Role in Climate Services, Adaptation, and Applications [PowerPoint Slides].Retrieved from https://www.legislative.noaa.gov/policybriefs/Climate%20Adaptation%20Briefing_Final022709.pdf

Konkel, F. (2014). The Details About the CIA's Deal with Amazon. *The Atlantic*. Retrieved from: https://www.theatlantic.com/technology/archive/2014/07/the-details-about-the-cias-deal-with-amazon/374632/

Kwon, D. (October 1, 2015). Are Europeans Better Than Americans at Forecasting Storms? Scientific American. Retrieved from: https://www.scientificamerican.com/article/are-europeans-better-than-americans-at-forecasting-storms1/

Lampland, M., & Star, S. (Eds.). (2009). *Standards and their stories: How quantifying, classifying, and formalizing practices shape everyday life*. Ithaca, NY: Cornell University Press.

Leone, D. A., Endlich, R. M., Petričeks, J., Collis, R. T. H., & Porter, J. R. (1989). Meteorological considerations used in planning the NEXRAD network. Bulletin of the American Meteorological Society, 70(1), 4-13.

Lessig, L. (2013). "Institutional corruption" defined. The Journal of Law, Medicine & Ethics, 41(3), 553-555.

Little, L. R., Parslow, J., Fay, G., Grafton, R. Q., Smith, A. D., Punt, A. E., & Tuck, G. N. (2014). Environmental derivatives, risk analysis, and conservation management. *Conservation Letters*, *7*(3), 196-207.

Longden, A. (2017). Review of meteorological market and emergence of the information economy. Presented at the Workshop on Improving the Socio-economic Impact of NWP Data, Reading, England, March 1, 2017. Retrieved from: https://www.ecmwf.int/sites/default/files/elibrary/2017/17105-state-meteorological-market_0.pdf

Lubchenco, J. (2012). NOAA Response to Towards Open Weather and Climate Services. Retrieved from: https://sab.noaa.gov/sites/SAB/Documents/SAB%20Working%20Groups/Standing%20Working%20Groups/EISWG/NOAA_Response/NOAA_response_to_SAB_recommendations_on_OWCS_final.pdf

Mandates and Drivers (2018). Retrieved from https://nrc.noaa.gov/CouncilProducts/ResearchPlans/5YearRDPlan/NOAA5YRPHome/Appendices/MandatesandDrivers.aspx

Massachusetts Institution of Technology (2019). *Lincoln Laboratory: About*. Retrieved from: https://www.ll.mit.edu/about.

Mettler, S. & SoRelle, M. (2014). "Policy Feedback," *Theories of the Policy Process*, 3rd ed. Edited by Paul Sabatier. Boulder, CO: Westview Press, 151-81.

Mettler, S. (2002). "Social Citizens of Separate Sovereignties: Governance in the New Deal Welfare State," in Sidney M. Milkis and Jerome M. Mileur, *The New Deal and the Triumph of Liberalism*. Amherst, MA: University of Massachusetts Press.

Mettler, S. (2010). "Reconstituting the submerged state: The challenges of social policy reform in the Obama era." *Perspectives on Politics* (8)3, 803-824.

Mettler, S. (2011). *The Submerged State: How invisible government policies undermine American democracy*. Chicago, IL: University of Chicago Press.

Morash, M. (1984). Establishment of a juvenile police record. *Criminology*, *22*(1), 97-111.

Morton, H. (2016). Use of Credit Information in Insurance 2016 Legislation. Retrieved from:

http://www.ncsl.org/research/financial-services-and-commerce/use-of-credit-information-in-insurance-2016-legislation.aspx

National Center for Atmospheric Research (2019). History: Meeting Challenges with Creativity. Retrieved from: https://ncar.ucar.edu/who-we-are/history

National Environmental Satellite, Data, and Information Service (2016). NOAA Observing System Integrated Analysis (NOSIA-II) Methodology Report https://www.nesdis.noaa.gov/sites/default/files/asset/document/noaa_tech_report_nesdis_147_nosia_ii_methodology_v1_95.pdf

National Information Consortium. (2015). Form 10-K 2014. Retrieved from Security and Exchange Commission website at: https://www.sec.gov/Archives/edgar/data/1065332/000115752315000683/a51038672.htm

National Oceanic and Atmospheric Administration (1991). The National Weather Service and the Private Weather Industry: A Public-Private Partnership, 56 Federal Register 1984, January 18, 1991.

National Oceanic and Atmospheric Administration (2004a). (Draft) Policy and Partnerships in the Provision of Weather, Water, Climate, and Related Environmental Information. Retrieved from https://web.archive.org/web/20061011060214/http://www.nws.noaa.gov/partnershippolicy/fairwxpolicy.htm

National Oceanic and Atmospheric Administration (2004b). (Final) Policy on Partnerships in the Provision of Environmental Information. Retrieved from https://web.archive.org/web/20061011060206/http://www.nws.noaa.gov/partnershippolicy/dec2004.htm

National Oceanic and Atmospheric Administration (2005). Comments on "Proposed Clarification to NOAA's Policy on Partnerships in the Provision of Environmental Information" Posted August 4, 2005. Retrieved from https://web.archive.org/web/20061014174911/http://www.nws.noaa.gov/partnershippolicy/ppclarificationcomments.pdf

National Oceanic and Atmospheric Administration (2006). NAO 216-112: Policy on Partnerships in the Provision of Environmental Information. Retrieved from http://www.noaa.gov/work-with-us/partnership-policy

National Oceanic and Atmospheric Administration (2010). Next Generation Strategic Plan. Retrieved from http://www.performance.noaa.gov/wp-content/uploads/NOAA_NGSP.pdf

National Oceanic and Atmospheric Administration (2012). *NOAA Research News: NOAA, Willis Re sign cooperative research agreement to improve hail damage estimates* (July 17, 2012). Retrieved from: https://research.noaa.gov/article/ArtMID/587/ArticleID/1407/NOAA-Willis-Re-sign-cooperative-research-agreement-to-improve-hail-damage-estimates

National Oceanic and Atmospheric Administration (2015). *L. Patents and Cooperative Research and Development Agreements (CRADAs) with industry (1i)*. NOAA National Severe Storms Laboratory Review 2015. Retrieved from: https://www.nssl.noaa.gov/about/events/review2015/support/L_Patents_and_CRADAs_with_industry.pdf

National Oceanic and Atmospheric Administration (2018a). NOAA RFI Cloud Data and Access Services: BDP Interim Lessons Learned-Executive Summary-Public. Retrieved from: https://www.fbo.gov/utils/view?id=9f985fa5ed4cfd40d8ad5f630a79c036

National Oceanic and Atmospheric Administration (2018b). NOAA's Request for Information (RFI) for Cloud and Data Access Services. Retrieved from: https://web.archive.org/web/20181214011950/https://www.fbo.gov/utils/view?id=7839b6770cd91b86d5c8818e78637b73

National Oceanic and Atmospheric Administration (n.d.). *GOES-R, Building Better Predictors*. Retrieved from: https://www.ncei.noaa.gov/news/goes-r-building-better-predictors

National Oceanic and Atmospheric Administration and National Aeronautics and Space Administration (n.d.). *User Systems: GOES Rebroadcast*. Retrieved from: https://www.goes-r.gov/users/grb.html

National Research Council (1995) National Weather Service Modernization Committee – Weather Panel: *Assessment of NEXRAD coverage and associated weather services.* National Academies Press. Retrieved from: https://www.nap.edu/catalog/9056/assessment-of-nexrad-coverage-and-associated-weather-services

National Research Council (1999). *Enhancing Access to NEXRAD Data - A Critical National Resource.* Washington, DC: The National Academies Press. https://doi.org/10.17226/9642.

National Research Council (2001). *A Climate Services Vision: First Steps Toward the Future*. Washington, DC: The National Academies Press.

National Research Council (2002). *Weather Radar Technology Beyond NEXRAD*. Washington, DC: The National Academies Press. Retrieved from: https://www.nap.edu/catalog/10394/weather-radar-technology-beyond-nexrad

National Research Council (2003). *Fair Weather: Effective Partnership in Weather and Climate Services*. Washington, DC: The National Academies Press. https://doi.org/10.17226/10610.

National Research Council, (2009). Informing Decisions in a Changing Climate. Washington, DC: The National Academies Press. https://doi.org/10.17226/12626.

National Research Council, (2012). Weather Services for the Nation: Becoming Second to None." Retrieved from https://www.nap.edu/catalog/13429/weather-services-for-the-nation-becoming-second-to-none

National Severe Storms Laboratory (2019). *Research Tools: NEXRAD*. Retrieved from https://www.nssl.noaa.gov/tools/radar/nexrad/

National Weather Service (2013). *What is Dual Polarization Radar and What Can it Do for Me?* PowerPoint Presentation. Retrieved from: https://www.weather.gov/media/lmk/soo/Dual_Pol_Overview.pdf

National Weather Service Radar Operations Center (2003). Registered Participants Attending the June 18, 2003 NWS Meeting to Discuss NWS Plans To Collect and Distribute WSR-88D Level II Data. Retrieved from: https://web.archive.org/web/20040520163519/http://www.roc.noaa.gov/NWS_Level_2/ResultsFrom061803Mtg/061803FinalAttendanceList.pdf

National Weather Service Radar Operations Center (2003a). *Additional Questions NOAA's National Weather Service (NWS) Has Received Since the July 25, 2003 Posting of Questions Based On The Public Meeting to Discuss the NWS Plans to Collect, Distribute, and Archive WSR-88D Level II Data Held At Silver Spring, MD on June 18, 2003 - In Addition, Questions Not Answered In The July 25 Posting Are Restated (As Of September 3, 2003).* Retrieved from: https://web.archive.org/web/20040520103409/http://www.roc.noaa.gov/NWS_Level_2/ResultsFrom061803Mtg/061803Mtg_Questions.pdf

National Weather Service Radar Operations Center (2003b). *Summary of NOAA's National Weather Service (NWS) Public Meeting To Discuss the NWS plans to collect, distribute, and archive WSR-88D Level II Data. Held At Silver Spring, MD on June 18, 2003 (As Of July 7, 2003).* Retrieved from: https://web.archive.org/web/20050308080029/http://www.roc.noaa.gov/NWS_Level_2/ResultsFrom061803Mtg/

National Weather Services Duties Act of 2005, S. 786, 109th Cong. (2005). Retrieved from https://www.congress.gov/109/bills/s786/BILLS-109s786is.pdf

Natural Capital (NOAA Performance, Risk, and Social Science Office (2018). http://www.performance.noaa.gov/natural-capital/

Neilley, P., Kyger, B., Ramamurthy, M. (2014). *Towards a Community Resource for High-Volume Data Processing near NCEP*. Presentation at the 94th meeting of the American Meteorological Society, Thursday, February 6, 2014 in Atlanta, Georgia. Retrieved from https://ams.confex.com/ams/94Annual/videogateway.cgi/id/26363?recordingid=26363

Nemes, V., La Nauze, A., & O'Neill, J. (2011). Contracts for environmental outcomes: the use of financial contracts in environmental markets. In *2011 Conference (55th), February 8-11, 2011, Melbourne, Australia* (No. 100576). Australian Agricultural and Resource Economics Society.

NEXRAD Joint System Program Office (1985). Next Generation Weather Radar Algorithm Report. Retrieved from: https://archive.org/details/nextgenerationwe00nexr

NEXRAD Joint System Program Office (1991). *NEXRAD Technical Requirements (NTR) R400-SP401A*. Retrieved from: https://www.roc.noaa.gov/wsr88d/publicdocs/ntr91.pdf

NEXRAD Technical Advisory Committee (1993). *Minutes of TAC meeting in Leesburg and Sterling, VA on December 6-7, 1993*. Retrieved from: https://web.archive.org/web/20001012073256fw_/http://www.osf.noaa.gov/app/tac/past_mtgs/MinDec93.htm

NEXRAD Technical Advisory Committee (1994). *Minutes of TAC meeting in Tulsa, Oklahoma, July 19-21, 1994*. Retrieved from: https://web.archive.org/web/20010731022235fw_/http://www.osf.noaa.gov/app/tac/past_mtgs/MinJul94.htm

Neustadt, Richard E. 1960. *Presidential Power and the Modern Presidents: The Politics of Leadership*. New York: Wiley.

NIC, INC. (2018). 2018 10-K Form. Retrieved from: https://www.sec.gov/Archives/edgar/data/1065332/000106533219000006/egov-123118x10k.htm#s274984F200DB5118B0C659362D007CD8

Nix, N. (2018). Google Drops Out of Pentagon's $10 Billion Cloud Competition. *Bloomberg.com*. Retrieved from: https://www.bloomberg.com/news/articles/2018-10-08/google-drops-out-of-pentagon-s-10-billion-cloud-competition

NOAA (2015). Cooperative Research & Development Agreement: CRADA Identification Number TT-000146. Retrieved from: https://content-drupal.noaa.gov/sites/default/files/atoms/files/NOAA_BDP_CRADA_final_draft.pdf

NOAA (2018). Cooperative Research and Development Agreements (CRADAs). Retrieved from: https://techpartnerships.noaa.gov/Partnerships-Licensing/CRADAs

NOAA (n.d). How to participate. Retrieved from: https://www.noaa.gov/big-data-project-how-to-participate

NOAA Cooperative Agreement Renewed (February 23, 2015). Retrieved from https://sciences.ncsu.edu/news/noaa-cooperative-agreement-renewed/

NOAA Cooperative Institutes (2018). Retrieved from https://ci.noaa.gov/CooperativeInstitutes.aspx

NOAA Data Catalog (2018). Retrieved from https://data.noaa.gov/dataset/

NOAA Observing Systems Council (NOSC) (2018). https://nosc.noaa.gov/

NOAA Organization (2018). Retrieved from http://www.noaa.gov/about/organization

NOAA Science Advisory Board - Working Group to Examine Advisory Options for Improving Communications among NOAA's Partners (Partnerships Group) (2008). Final Report: Advisory Options for Improving Communications among NOAA's Partners. Retrieved from: https://web.archive.org/web/20100527091815/http://www.sab.noaa.gov/Reports/2008/NOAA_SAB_Partnerships_Report_Sep08_FINALtoNOAA.pdf

NOAA Science Advisory Board (2009b). *EISWG Terms of Reference*. Retrieved from https://web.archive.org/web/20100527105840/http://www.sab.noaa.gov/Working_Groups/standing/docs/2009/NOAA_SAB_EISWG_ToR_FINAL_Apr2009.pdf

NOAA Science Advisory Board (SAB) - Environmental Information Services Working Group (EISWG). *Meeting Summary*, November 18 – 20, 2009, Silver Spring, Maryland. Retrieved from https://sab.noaa.gov/sites/SAB/Documents/SAB%20Working%20Groups/Standing%20Working%20Groups/EISWG/2009_11_EISWG_MeetingSummary.pdf

NOAA Science Advisory Board (SAB) (2014b) - Environmental Information Services Working Group (EISWG) Meeting, (handout) *Latest on CRADA and Open Environmental Information Services*, Retrieved from: ftp://ftp.oar.noaa.gov/SAB/EISWG/2014_Dec/NOAA_SAB_EISWG_Dec2014_LatestonCRADAandOpenEnvironmentalInformationServices.pdf

NOAA Science Advisory Board (SAB) (2014c) - Environmental Information Services Working Group (EISWG) Meeting, *(handout) Open Weather and Climate Services*, Retrieved from: ftp://ftp.oar.noaa.gov/SAB/EISWG/2014_Dec/NOAA_SAB_EISWG_Dec2014_Neilley_OpenWeatherandClimateServices.pdf

NOAA Science Advisory Board (SAB) (2014d) - Environmental Information Services Working Group (EISWG) Meeting, *Status of Open Environmental Information Services (Open EIS)*, Retrieved from: ftp://ftp.oar.noaa.gov/SAB/EISWG/2014_Dec/NOAA_SAB_EISWG_Dec2014_Johnson

_OpenEIS.pdf

NOAA Science Advisory Board (SAB), (2014) - Environmental Information Services Working Group (EISWG), *Agenda - Twelfth Meeting of the NOAA Science Advisory Board's Environmental Information Services Working Group*, December 2-3, 2014, Silver Spring, Maryland. Retrieved from: ftp://ftp.oar.noaa.gov/SAB/EISWG/2014_Dec/NOAA_SAB_EISWG_Dec2014_Agenda_FINAL.pdf

NOAA Science Advisory Board, 2010. *Members of the Environmental Information Services Working Group (EISWG).* Retrieved from https://web.archive.org/web/20100527090533/http://www.sab.noaa.gov/Working_Groups/standing/environmental/members.html

NOAA Science Advisory Board, Environmental Information Services Working Group (EISWG) (2011). Towards Open Weather and Climate Services. Retrieved from: https://www.sab.noaa.gov/sites/SAB/Documents/SAB%20Working%20Groups/Standing%20Working%20Groups/EISWG/Reports/Towards-Open-Weather-and-Climate-Services-report-and-transmittal_12_23_11.pdf

North Carolina Institute for Climate Studies, (2012). Alternative Climate Normals Workshop. Retrieved from: https://ncics.org/news/events/alternative-climate-normals-workshop/

NSTC (National Science and Technology Council) (2014). *National Plan for Civil Earth Observations.* Retrieved from: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/NSTC/2014_national_plan_for_civil_earth_observations.pdf

Office of Management and Budget (OMB) (1976). *Circular No. A-109, Major Systems Acquisitions.* Retrieved from: https://web.archive.org/web/20170120233952/https://www.emcbc.doe.gov/pmo/supporting_files/omb_circular_a_109.pdf

Office of Management and Budget (OMB) (1993). Circular Number A-25, *Revised. User Charges.* Retrieved from: http://www.whitehouse.gov/omb/circulars_a025

Office of Management and Budget (OMB) (2015). Memorandum M-16-01 Incorporating Ecosystem Services into Federal Decision Making. Retrieved from https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2016/m-16-01.pdf

Office of Management and Budget (OMB) (2016). *Circular No. A-130, Managing Information as a Strategic Resource.* Retrieved from: https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/OMB/circulars/a130/a130revised.pdf

Office of Management and Budget (OMB) (2020a). *Federal Data Strategy: Leveraging Data as a Strategic Asset.* Retrieved from: https://strategy.data.gov/

Office of Management and Budget (OMB) (2020b). *Practices for Leveraging Data as a Strategic Asset*. Retrieved from: https://strategy.data.gov/assets/docs/federal-data-strategy-practices.pdf

Oreskes, N. (2003). A context of motivation: US Navy oceanographic research and the discovery of sea-floor hydrothermal vents. Social Studies of Science, 33(5), 697-742.

Overman, E. & Cahill, A. (1990). Information policy: A study of values in the policy process. *Review of Policy Research*, *9*(4), 803-818.

Patashnik, E. & Zelizer, J. (2009). "When policy does not remake politics: The limits of policy feedback." In *Annual Meeting of the American Political Science Association in Toronto*. 2009.

Peled, A. (2011). When transparency and collaboration collide: The USA open data program. *Journal of the American society for information science and technology*, *62*(11), 2085-2094.

Pettifer, R. (2015). The development of the commercial weather services market in Europe: 1970–2012. *Meteorological Applications*, 22(3), 419-424.

Pfeffer, J., & Salancik, G. R. (2003). *The external control of organizations: A resource dependence perspective*. Stanford University Press.

Pierson, P. (1993). "When effect becomes cause: Policy feedback and political change." *World Politics* 45(4): 595-628.

Pike, A., & Pollard, J. (2010). Economic geographies of financialization. *Economic geography*, *86*(1), 29-51.

Privacy Act of 1974, Amended System of Records, 82 Fed. Reg. 8 (January 12, 2017). Federal Register: The Daily Journal of the United States.

Privacy Act Statement (2018). Retrieved from https://www.ncdc.noaa.gov/about-ncdc/privacy/privacy_act_statement

Privacy Impact Assessment for the National Climatic Data Center Local Area Network (NOAA5009) (2018). Retrieved from: http://www.osec.doc.gov/opog/privacy/NOAA%20PIAs/NOAA5009_PIA_SAOP_Approved.pdf

Privacy Threshold Analysis for the National Centers for Environmental Information – NC NOAA5009 (2018). Retrieved from: http://www.osec.doc.gov/opog/privacy/NOAA%20PIAs/NOAA5009_PTA_2018.pdf

Quandl (2018). https://www.quandl.com/about

Radar Operations Center (2009). *Software Development and Maintenance Guide (WPI0036)*. Retrieved from: https://www.roc.noaa.gov/wsr88d/PublicDocs/ENG/WPI0036.pdf

Radar Operations Center (2009a). *Software Development and Maintenance Process. Radar Operations Center Software Engineering*. Retrieved from: https://www.roc.noaa.gov/WSR88D/PublicDocs/SDMP.pps

Radar Operations Center (2015). *NOAA/National Weather Service Radar Functional Requirements*. Retrieved from: https://www.roc.noaa.gov/WSR88D/PublicDocs/NOAA_Radar_Functional_Requirements_Final_Sept%202015.pdf

Radar Operations Center (2019). *Dual-polarization Technology Overview*. Retrieved from: https://training.weather.gov/wdtd/courses/dualpol/Outreach/non-mets-intro/presentation.html

Randalls, S. (2010). Weather profits: Weather derivatives and the commercialization of meteorology. *Social Studies of Science*, *40*(5), 705-730.

Raytheon (2019). *Advanced Weather Interactive Processing System (AWIPS).* Retrieved from: https://www.raytheon.com/capabilities/products/awips

Regional Climate Centers (2018). Retrieved from: https://www.ncdc.noaa.gov/customer-support/partnerships/regional-climate-centers

Roberts, A. (2006). *Blacked out: Government secrecy in the information age*. Cambridge, UK: Cambridge University Press.

Ruijer, E., Détienne, F., Baker, M., Groff, J., & Meijer, A. J. (2019). The Politics of Open Government Data: Understanding Organizational Responses to Pressure for More Transparency. The American Review of Public Administration, 0275074019888065.

Ruppert, E. (2015), “Doing the Transparent State: Open government data as performance indicators,” in *The World of Indicators: The Making of Governmental Knowledge through Quantification*. Rottenburg, R., Merry, S. E., Park, S. J., & Mugler, J. (Eds.) pp. 51–78. Cambridge, UK: Cambridge University Press.

Safeguarding Information (2018). http://www.osec.doc.gov/opog/privacy/PII_BII.html

Samenow, J. (2017) *The best forecasts for Hurricane Irma came from a computer model few people know about*. Washington Post. Retrieved from https://www.washingtonpost.com/news/capital-weather-gang/wp/2017/09/18/the-best-forecasts-for-hurricane-irma-came-from-a-computer-model-few-people-know-about/?noredirect=on&utm_term=.1824245dcd49

Sankey, M. (2015). The MVR Access and Decoder Digest: 2015/2016. BRB Publications, Inc. Tempe, Arizona.

Schattschneider, E. (1935). *Politics, Pressure, and the Tariff*. New York, NY: Prentice-Hall.

Schellenberg, T. R. (1956). *Modern Archives: Principles and Techniques*. Chicago, IL: University of Chicago Press.

Schwartz, J. & Stelter, B. (2012, July 3). Fans Howl After Weather Site Buys Out Rival. The New York Times. Retrieved from: https://www.nytimes.com/2012/07/04/us/as-weather-channel-buys-weather-underground-fans-fear-change.html

Shaban, B, et. al (2017 January 24). SF Restaurants With Poor Health Ratings Can Now Buy Opportunity for New Scores. NBC Bay Area. Retrieved from: https://www.nbcbayarea.com/news/local/SF-Restaurants-With-Poor-Health-Ratings-Can-Now-Buy-Opportunity-for-New-Scores-411721205.html

Simon, P. A., Leslie, P., Run, G., Jin, G. Z., Reporter, R., Aguirre, A., & Fielding, J. E. (2005). Impact of restaurant hygiene grade cards on foodborne-disease hospitalizations in Los Angeles County. *Journal of Environmental Health*, 67(7).

Skocpol, T. (1995). *Protecting Soldiers and Mothers*. Cambridge, MA: Harvard University Press.

Smith, D. (1974). The Social Construction of Documentary Reality. *Sociological Inquiry*, 44(4), 257-268.

Smith, M. (2013). Extraordinary Tornadoes and Preparedness Best Practices [PowerPoint slides]. Retreived from: https://enterprisesolutions.accuweather.com/assets/documents/Extraordinary_Tornadoes_and_Preparedness_Best_Practices.pdf

Solomon, S, Dole, R., Feely, R., Held, I., Higgins, W., Payne, J., Shea, E., Varanasi, E. Westley, M. (2009). *A Vision for Climate Services in NOAA*. Retrieved from: https://d32ogoqmya1dw8.cloudfront.net/files/cln/vision_climate_services_noaa.pdf

Sousounis, P. (2006). *Real Time Aviation Forecasting at WSI.* Presented at the 22nd International Conference on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology. Held January 27 - February 03, 2006 in Atlanta, GA.

Star, S. & Griesemer, J. (1989). Institutional ecology, translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, *19*(3), 387-420.

Star, S. (1999). The ethnography of infrastructure. *American behavioral scientist*, *43*(3), 377-391.

Stokes, D. E. (1997) Pasteur's Quadrant: Basic Science and Technological Innovation. Washington, D.C. Brookings Institution Press.

Success Stories on User Engagement (2018). https://www.ncdc.noaa.gov/success

Sullivan, S. (2013). Banking nature? The spectacular financialisation of environmental conservation. *Antipode*, *45*(1), 198-217.

Sunlight Foundation (2016). *Open Data Policies at Work*. Retrieved from: https://sunlightfoundation.com/policy/opendatamap/.

Tupasela, A., Snell, K., & Tarkkala, H. (2020). The Nordic data imaginary. Big Data & Society, 7(1), 2053951720907107.

Texas Comptroller of Public Accounts (2015). *Sources of Revenue: A History of State Taxes and Fees in Texas 1972 to 2014*. Austin, TX. Retrieved from http://www.texastransparency.org/State_Finance/Revenue/Sources/2015/Sources_of_Revenue_2015.pdf

Texas Comptroller of Public Accounts (2019). *Sources of Revenue: A History of State Taxes and Fees in Texas 1972 to 2018*. Austin, TX. Retrieved from: https://comptroller.texas.gov/transparency/revenue/docs/96-571.pdf

The DIAS: User-friendly Access to Copernicus Data and Information (n.d). Retrieved from: https://www.copernicus.eu/sites/default/files/Copernicus_DIAS_Factsheet_June2018.pdf

The Weather Company (2015). IBM Plans to Acquire The Weather Company's Product and Technology Businesses; Extends Power of Watson to the Internet of Things (October 28, 2015). Retrieved from: https://business.weather.com/news/ibm-plans-to-acquire-the-weather-companys-product-and-technology-businesses-extends-power-of-watson-to-the-internet-of-things

TPIO Project Spotlight: NOAA Observing System Integration Analysis (NOSIA-II) (2018). Retrieved from https://nosc.noaa.gov/tpio/main/nosia_main.html

Trace, C. (2002). What is Recorded is Never Simply 'What Happened': Record Keeping in Modern Organizational Culture. *Archival Science*, *2*(1), 137–159.

U.S. Department of Commerce (2013). Fact Sheet: U.S. Department of Commerce "Open for Business Agenda." Retrieved from: https://2010-2014.commerce.gov/news/fact-sheets/2013/11/14/fact-sheet-us-department-commerce-open-business-agenda.html

U.S. Department of Commerce, National Oceanic and Atmospheric Administration (2014a). *Big Data RFI from NOAA*, February 21, 2014. Retrieved from: https://www.fbo.gov/index?s=opportunity&mode=form&tab=core&id=d9844cb78b4527fb11a6ac6d2b80a742&_cview=0

U.S. Department of Commerce, National Oceanic and Atmospheric Administration (2014b). *Big Data RFI from NOAA – Amendment 1*, March 21, 2014. Retrieved from: https://www.fbo.gov/index?s=opportunity&mode=form&tab=core&id=d0a8d9a279c69aeac9108507ff32f22a&_cview=0

U.S. Department of Commerce, National Oceanic and Atmospheric Administration (2014c). *Big Data RFI from NOAA – Special Notice*, September 26, 2014. Retrieved from: https://www.fbo.gov/index?s=opportunity&mode=form&tab=core&id=5c510b609df6e95250676b7101cf4ddc&_cview=0

U.S. Department of Commerce, National Oceanic and Atmospheric Administration (2014d). *Big Data RFI from NOAA – Modification / Amendment*, October 3, 2014. Retrieved from: https://www.fbo.gov/index?s=opportunity&mode=form&tab=core&id=cdbfd2f6b096dfe93aecae44b67fcc40&_cview=0

U.S. Department of Commerce, National Oceanic and Atmospheric Administration (2014e). *Big Data RFI from NOAA – Attachment B: Draft Big Data Statement of Objectives*, October 3, 2014. Retrieved from: https://www.fbo.gov/utils/view?id=4104ec928673e8f6b5cbd741a833b406

U.S. Department of Commerce, National Oceanic and Atmospheric Administration (2014f). *Big Data RFI from NOAA – Modification: Documents from Big Data Industry Day on 10/17/2014 2014-10-17 BDP Industry Day Presentation*, October 28, 2014. Retrieved from: https://www.fbo.gov/utils/view?id=797d5e47961c5c9ab33b64c65dc6eb34

U.S. Department of Commerce, National Oceanic and Atmospheric Administration (2014g). *Big Data RFI from NOAA – Modification: Documents from Big Data Industry Day on 10/17/2014, October 28, 2014*. Retrieved from: https://www.fbo.gov/index?s=opportunity&mode=form&tab=core&id=d85395e880589db0f35edef39e71314b&_cview=0

U.S. Department of Commerce, National Oceanic and Atmospheric Administration (2014h). *Big Data RFI from NOAA – Modification: Documents from Big Data Industry Day on 10/17/2014: 2014-10-17 Questions and Answers*, October 28, 2014. Retrieved from: https://www.fbo.gov/utils/view?id=460e882195422445be6a24d349e6d782

U.S. Department of Commerce, National Oceanic and Atmospheric Administration (2015). *Big Data RFI from NOAA – Modification: For Immediate Release, April 21, 2015*. Retrieved from: https://www.fbo.gov/index?s=opportunity&mode=form&tab=core&id=f54565059caaa2e97722f49bebc96477&_cview=0

U.S. Department of Commerce, Public Affairs Office (2015). U.S. Secretary of Commerce Penny Pritzker Announces New Collaboration to Unleash the Power of NOAA's Data [Press release], April 21, 2015. Retrieved from: https://web.archive.org/web/20150423235525/https://www.commerce.gov/news/press-releases/2015/04/us-secretary-commerce-penny-pritzker-announces-new-collaboration-unleash

U.S. Food and Drug Administration (2016). FDA Investigates Multistate Outbreak of E. coli O26 Infections Linked to Chipotle Mexican Grill Restaurants. February 1, 2016. Retrieved from: https://www.fda.gov/food/outbreaks-foodborne-illness/fda-investigates-multistate-outbreak-e-coli-o26-infections-linked-chipotle-mexican-grill-restaurants

U.S. Food and Drug Administration (2018). Adoption of the FDA Food Code by State and Territorial Agencies Responsible for the Oversight of Restaurants and Retail Food Stores. Retrieved from: https://www.fda.gov/media/107543/download

U.S. Food and Drug Administration (2019). FDA Food Code. Retrieved from https://www.fda.gov/food/retail-food-protection/fda-food-code

Unidata (2014) Open Weather & Climate Services “Servers @ NCEP” Experiment Project. Retrieved from: https://web.archive.org/web/20140328095752/http://www.unidata.ucar.edu/projects/index.html#owcs

Unidata (2019). LDM Fact Sheet. Retrieved from: https://www.unidata.ucar.edu/software/ldm/ldm-current/factsheet.html

United States Department of Commerce (2014). *Fostering Innovation, Creating Jobs, Driving Better Decisions: The Value of Government Data*. Washington, DC. Retrieved from: http://www.esa.doc.gov/sites/default/files/revisedfosteringinnovationcreatingjobsdrivingbetterdecisions-thevalueofgovernmentdata.pdf

United States Department of Justice. (2012). *Survey of State Criminal History Information Systems 2012*. Retrieved from https://www.ncjrs.gov/pdffiles1/bjs/grants/244563.pdf

United States Federal Trade Commission (2014). *Data Brokers: A call for transparency and accountability*. Washington, DC. Retrieved from: https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf

University Corporation for Atmospheric Research (UCAR), (2017). NCAR/UCAR Research Applications Laboratory 2017 Annual Report. Retrieved from: https://nar.ucar.edu/sites/default/files/labs/ral/2017/2017%20RAL%20Annual%20Report.pdf

University of Oklahoma (2001). *National Stakeholder's Workshop on the Acquisition, Use, and Archival of WSR-88D Level II Data.* February 14-16, 2001. Boulder, Colorado. Retrieved from: http://kkd.ou.edu/FINAL%20Agenda.pdf

University of Oklahoma (2002). *Summary of the Second WSR-88D Level II-Data Stakeholders Workshop*. September 26-27, 2002. Norman, Oklahoma. Retrieved from: http://kkd.ou.edu/Summary%20of%202002%20CRAFT%20Workshop.pdf

University of Oklahoma (2003). *Call for Private Sector Participation in the Collaborative Radar Acquisition Field Test (CRAFT): The Distribution of WSR-88D (NEXRAD) Radar Level II Data in Near Real Time.* Retrieved from: http://kkd.ou.edu/private.pdf

University of Oklahoma (2019). *A History of CAPS…* Retrieved from: http://www.caps.ou.edu/about_caps.php

Urhart, M. (2009) NOAA Science Advisory Board: Background and Relationship to its Working Groups A Presentation to the SAB Environmental Information Services Working Group, presentation made to the Environmental Information Services Working Group, November 19, 2009, Silver Spring, MD ftp://ftp.oar.noaa.gov/SAB/EISWG/2009_Nov/NOAA_SAB_EISWG_Nov2009_Uhart.pdf

User Engagement (2018). Cooperative Institute for Climate and Satellites – North Carolina. https://ncics.org/portfolio/outreach-engagement-3/ncei-user-engagement/

Van Maanen, J., & Pentland, B. (1994). Cops and Auditors: The Rhetoric of Records. In Sitkin, S. & Bies, R. (Eds.), *The Legalistic Organization* (pp. 53-90). Thousand Oaks, CA: Sage Publications, Inc.

Veispak, A. (2017) Copernicus Data and Access Information Services (DIAS). Presented at the European Big Data Value Forum, Earth Observation Big Data Workshop, November 23, 2017, Versailles, France. Retrieved from: https://www.european-big-data-value-forum.eu/wp-content/uploads/2018/01/AndreasVeispak-Copernicus-DIAS-EBDVF17.pdf

Venuti, F. (2017). *ECMWF Data Services Future*. Presented at the Workshop on Improving the Socio-economic Impact of NWP Data, Reading, England, March 1, 2017. Retrieved from: https://www.ecmwf.int/sites/default/files/elibrary/2017/17107-future-data-services.pdf

Venuti, F., Andersson, E., Modigliani, U., Pidduck, E., Gutierrez, M., Blanchonnet, H., and Plumridge, M. (2018) The Evolution of the European Centre for Medium-Range Weather Forecasts (ECMWF) Data Services. [poster] Presented at the 98th Annual Meeting of the American Meteorological Society, January 9, 2018 in Austin, Texas. Retrieved from: https://ams.confex.com/ams/98Annual/webprogram/Handout/Paper334268/2018-01-09_AMS_Poster_ECMWF_Future_Data_Services.pdf

Vickery, G. (2011). *Review of recent studies on PSI re-use and related market developments.* Europe's Information Society Thematic Portal. Paris: European Commission. Retrieved from: http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=1093

Vingione, G. (2018) Data and Information Access Services. Presented at the Joint Ad-hoc Meeting of the IGS and the Collaborative GS, February 7, 2018, Brussels, Belgium. Retrieved from: http://www.imodi.info/wp-content/uploads/2018/03/05_DIAS-SERCO-V6.pdf

Weymanapri, J, Clancy, K. (1989). Next Generation Weather Radar (NEXRAD) Initial Operational Test and Evaluation. Phase 2. (IOT&E(2)). Air Force Operational Test and Evaluation Center Kirtland Air Force Base, New Mexico. Retrieved from: https://apps.dtic.mil/dtic/tr/fulltext/u2/a218770.pdf

Whiton, R. C., Smith, P. L., Bigler, S. G., Wilk, K. E., & Harbuck, A. C. (1998). History of operational use of weather radar by US weather services. Part I: The pre-NEXRAD era. *Weather and Forecasting, 13(2)*, 219-243.

WIBW (2018). WIBW News Evening Weather Forecast, November 22, 2018. Retrieved from: https://www.wibw.com/video?vid=501045151

Wilhelm, T. (2016. Performance, Risk & Social Science Office, OCFO: Update to the Deputy Under Secretary for Operations (May 26, 2016). Retrieved from: http://www.performance.noaa.gov/wp-content/uploads/Tony-PRSS-Overview-June-2016.pptx

Yelp (2013 January 17). Mayor Ed Lee Announces Open Data Partnership with Yelp To Offer Restaurant Health Inspection Scores To Improve Public Health, Transparency [Press Release]. Retrieved from: https://www.prnewswire.com/news-releases/mayor-ed-lee-announces-open-data-partnership-with-yelp-to-offer-restaurant-health-inspection-scores-to-improve-public-health-transparency-187299731.html

Yelp (2019). *Local Inspector Value-Entry Specification*. Retrieved from: https://www.yelp.com/healthscores

Yelp (2019b). *Local Inspector Value-Entry Specification (LIVES). Current Municipality Feeds*. Retrieved from: https://www.yelp.com/healthscores/feeds

Zazulia, N. (October 12, 2018) Panasonic Avionics Sells Weather Business to FLYHT. Aviation Today. Retrieved from: https://www.aviationtoday.com/2018/10/12/panasonic-avionics-sells-weather-business-flyht/