Specify for Archaeology: A Proposed Data Model for Archaeological Collection Database Management

By

Theresa M. Miller

Submitted to the graduate degree program in Anthropology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Arts.

Chariperson: ______________________________

Dr. John W. Hoopes, Anthropology

Committee Members: ______________________________

Dr. James H. Beach, Director Specify Software Program

____

Dr. Ivana Radovanovic, Anthropology

____

Dr. Dixie West, Anthropology, Kansas State University

Date Defended: December 7, 2012

The Thesis Committee for Theresa M. Miller

certifies that this is the approved version of the following thesis:

Specify for Archaeology:

A Proposed Data Model for Archaeological Collection Database Management

Chairperson Dr. John W. Hoopes

December 7, 2012

Abstract

Archaeological databases are one of the most significant research resources for archaeologists. The databases currently available commercially were not designed for archaeological collections and are not optimal for documentation of archaeological data. Specify Software has created software that effectively documents the collection data of natural history collections for over 400 collections world-wide. This software could also be effective for archaeological collections as well if it were adapted for archaeological collections. This thesis examines how the software could be adapted in order to proficiently document archaeological data as well. In order to ascertain how Specify must be adapted, I determined the database requirements of an archaeological collection by reviewing excavation procedures and interviewing archaeologists about their current database systems. I examine the differences between database requirements for biological and archaeological collections and define the changes that must be made to the current Specify data model. I propose a new data model that will efficiently document archaeological data. I use raw archaeological data to evaluate how the proposed data model will work with archeological data.

Acknowledgements

First, I would like to thank my advisor, John Hoopes. He has worked with me and encouraged me these past years to try to produce the best possible thesis. I have not been the traditional student. My full-time job, moving around the country, and the birth of my baby boy, Benjamin, have definitely been challenges to my work and Dr. Hoopes has worked with me through all of it.

I would also like to my committee members who have given their time and guidance to make this thesis happen. I would especially like to thank Dixie and Jim who got me involved in this research in the first place. It was their original collaboration that produced an interest in this thesis. Their expertise and experiences have been vital to my research, not to mention bringing me with them to interviews and Dixie's workshop, and Dixie giving me my first chance to present at a conference.

The interviews I conducted have been an invaluable part of my research. I would like to thank all of the people who gave me their time and access to their resources including: James Beach, Andrew Bentley, Luis Jaime Castillo Butters, Kathy Coggins, Una Farrell, Christine Garst, Megan Innes, Liz Lieth, Laura Halverson Monahan, John Miller, Dawn Planas, Kerrie-Ann Shannon, Stephanie Teasley, Misty Tilson, James Whitney, Josh Wisniewski, and especially Mary Adair who not only let me pick her brain repeatedly (poor woman), but gave me access to some of her data as well. Mary has been very supportive, offering me any assistance I may need from her. James Beach and Rod Spears at Specify have been beyond invaluable offering me guidance, asking me questions to guide my research, and helping me to understand data models and how data is/can be related.

I must also thank the NSF who funded not only Dixie's project, which first started this thesis but also funds Specify Software, the employees of which I have worked with while writing this thesis and for which I have worked for the past several years.

I would also like to thank Luis Jaime Castillo Butters and Julio Rucabado-Yong who gave me such amazing archaeological experiences. The majority of my archaeological experience has occurred under their guidance. Both of them have helped shape me as an archaeologist and piqued my interest even more than I had ever been interested before in the Moche.

I would also like to thank Le-Thu, Judy Ross, and Jack Hofman who have helped me navigate the graduate school requirements.

Finally, I would like to thank my family who have been so very encouraging of me (i.e. pestering me to) completing my thesis. I would especially like to thank my husband who has had to deal with me putting off other responsibilities so I would work on completing my thesis. And, Benjamin, who just smiles at me or hugs me and puts everything into perspective and for whom I want to be a better person.

Table of Contents

Table of Figures

Table of Tables

Chapter 1 – Introduction

Archaeology offers a unique insight into the history of human culture and environment, but only if the information collected can be effectively represented and recorded. To this end, many attempts have been made toward the development of effective research information systems for archaeological collections. However, for various reasons, all of these database design and development initiatives have fallen short of being sustainable and successful. The Specify Software Project at the University of Kansas has the capacity to offer database software for archaeological collections that is as successful as its management software for biological collections. With the addition of several key data concepts, Specify can become an effective tool for managing archaeological collections.

Archaeology as a science of description, analysis, and interpretation is not isolated to field excavation activities. It is a process that begins before the archaeologist begins to collect and continues after an excavation is completed. Archaeologists strive to preserve as much information about the objects and other materials they collect as possible. To accomplish this goal, they must logically and thoroughly document and store objects and associated information. Both artifacts and the data associated with them must be readily accessible to researchers and students as well as to individuals who wish to review, consult, and reanalyze material in the future.

While antiquarian collections, collected by amateurs who amass objects for their aesthetic appeal, may be created solely for the intrinsic value of the object, systematic archaeological collections with accompanying documentation are created to produce and preserve information for the purposes of research and interpretation (Sullivan and Childs 2003). Scientists document objects in a systematic fashion. How this is achieved depends on the type of collection. A database created for a biological collection may not be sufficient to adequately document the vital information pertaining to objects collected in archaeological research.

Archaeological projects often include collaborators from a variety of disciplines, including biologists, geologists, paleontologists, architects, zoologists, and chemists. In 2006, Dr. Dixie West (University of Kansas) encountered a problem resulting from this multidisciplinary research. From 2005 to 2008, she directed a collaborative project in the Andreanof Island group of the Aleutian Islands, Alaska beginning with Adak Island. The project was an investigation of the connection between the environment and human settlement in the Central Aleutians during the Holocene. It incorporated the expertise of zooarchaeologists, historical ecologists, tephrachronologists, isotope experts, geomorphologists, microbiologists, and bioinformatics workers, in addition to archaeologists.

Multidisciplinary projects such as this, while presenting the advantage of more comprehensive understanding of the materials, can present archaeologists with significant challenges to capturing and recording all of the information documented in the course of an archaeological project by a variety of multidisciplinary experts. Existing database systems designed specifically for archaeologists currently lack the depth and breadth required to accomplish this goal. Other commercially available database systems not designed with these types of projects in mind are neither user-friendly nor do they offer the capacity to enter and retrieve the complex array of data generated in multidisciplinary studies.

In 2012, available collection database systems such as Specify, KE EMu, and Biota only allow for the information about collection objects from one academic discipline to be recorded in the database. Some systems partition the information they accommodate into sub-disciplines, for example, splitting biology into entomology, marine mammalogy, terrestrial mammalogy, botany, etc. This database design approach, which logically decomposes specimen or artifact data management into a single, discrete discipline with highly-constrained data types would, if used for archaeology, handle the information from one aspect of archaeological data to be recorded, and only part of the aspect at that. For example, basic object information may be recorded for a bone awl but nothing about the context in which it was discovered nor about the feature it came from (if it was discovered in a feature), nor much information about the taxonomy of the bone's source.

In an attempt to address this problem, West turned to the Specify Software Project (here forth referred to as Specify) at the University of Kansas, a program funded by the Approaches in Biological Informatics Program of the U.S. National Science Foundation. She began looking into the requirements for helping the existing Specify database system, initially developed use in biological collections of natural history museums, to evolve and meet the needs of a multidisciplinary archaeological project. West considered Specify to have the potential to effectively record all of the data that were essential to accomplishing the goals of her archaeological project. However, because Specify was developed for biological rather than archaeological collection description and management, there were problems satisfying the needs of archaeological data.

Substantial differences exist in the requirements of database systems for these distinct kinds of collections. These were primarily in relation to data concepts unique to archaeology and their representation in a database logical design or "schema". They presented unique challenges in conception and execution of database design. The central question addressed in this thesis is: Can Specify, a federally-funded, open source, database management system initially developed for biological collections, be successfully modified to accommodate the information management needs of archaeological collections?

Specify, designed to operate on Windows, Mac OS X, and Linux operating systems, was developed in the Informatics Department of the Museum of Natural History at the University of Kansas and as of September 2012, is used by over 247 natural history museums worldwide (www.specifysoftware.org). Licensed under a General Public License, Specify represents the continuation of the earlier MUSE Software Project, an earlier database system for biological collections developed in 1987. Under project director Dr. James Beach, Specify developed into a highly usable collection management database for biological collections. Its capabilities were implemented not only to manage collections; but also with an eye towards facilitating collection research efforts.

Specify permits research to be conducted outside of the facility housing the collection by providing a web server interface for controlled access by anonymous public users over the web. Database accessibility over the Internet is an invaluable asset to researchers and it opens up the collections to a larger public with controls for privacy and security. Easy accessibility promotes scientific collaboration that would otherwise be difficult because collections may otherwise be too distant or difficult to access in traditional ways. For example, many Arctic or Peruvian museums tend to be physically sequestered to a degree from the rest of the scientific community making them difficult to visit in person. Accessibility is also important for the functioning of small museums, many of which contain archaeological collections. Archaeological collections are often kept in smaller museums in geographical locations close to the sites where materials were excavated. They can also be stored in multiple facilities. The result can often be a collection pertinent to a specific site or region that is distributed among a series of disconnected, unrelated museums in different geographic locations. Increased accessibility allows for research on issues including tracing migration population patterns and cultural influence from one group on another, or studying the introduction of specific artifacts. Data integration, whether actual or virtual, can assist an investigator in forming datasets that provide a broader picture of the cultural and biological development of a region and the interaction of the two.

Specify was particularly appealing for this project because it is open-source “freeware”—free, widely available software that can be readily modified by different users--that has a user-friendly interface and can be highly customized for the needs of specific users. It allows for adaptation of forms to accommodate the user and the enforcement or restriction of terminology for institutionally consistent data entry. It also supports quick, full-text searches and more structured database queries on relevant search fields. A significant advantage for this thesis project was that Beach, who has been working with this database system for seventeen years, was available to provide assistance and expert insight into the functioning of Specify and with its modification for archaeological collections.

At the same time West was looking for an archaeologically-focused collection management software system that would be able to comprehensively manage her field research data, Beach was seeking to broaden the system to include other classes of collections. He recognized that archaeological collections were an area in which Specify could make a significant intellectual contribution. However, Beach had limited experience with archaeological collections. Completion of this project required combination of the of his informatics expertise with the research experience and collection management needs of West in order to conceptualize an information model for adapting Specify for use with archaeological collections.

For effective collection data management, adapting a database system from one discipline to another must involve much more than simply changing the names of database tables and fields to accommodate the differences in naming and object description between biological and archaeological collections. An adapted or modified system must be modified to accommodate the unique intellectual concepts and semantics of archaeological research. Beach needed to investigate and understand the differences in these requirements between the two kinds of databases and determine how to reconcile those conceptual differences with Specify's biological collection management database system before programmers could design a version of Specify for archaeological collection data.

To this end, Beach hired me to undertake an investigation of how Specify needed to be modified. Before being hired, I served as a software quality assurance tester for Specify. I was familiar with Specify's capabilities and database design, the ways in which existing users (in biological collections around the world) used the database, and the users' requirements for database functions. As a Master's student in archaeology at the University of Kansas with excavation experience in both the Netherlands and Peru, I was familiar with conceptual and practical issues associated with archaeological data.

This thesis represents the first of a multistage process that will result in the creation of an initial analysis and information domain: **Specify for Archaeology**. The research presented in this thesis explores the specific needs of archaeological research collection data management and the differences between archaeological and biological collection digital database requirements. Based on the findings of the investigation, a conceptual information model for archaeology was created. The second part of this thesis explains specifically how the model can be implemented.

Chapter 2 - Review

Chapter 2 examines the pattern in the discussion and development of database use in archaeology and briefly reviews significant literary contributions on the subject. The investigation began by examining the history of database design and by reviewing similarities and differences between biological and archeological collections. Database development has gone through a cyclical pattern of interest resulting in multi-year crests and troughs of activity and systems development. The cycle is driven by ever-changing research interests in the academic community, which in turn are influenced by shifts in technology. With significant advancement in technology, new potential capabilities for database use in archaeology emerge, encouraging debate and discussion of how those capabilities might best be achieved. Modern database systems used for archeology allow for the bulk of relevant scholarly information to be recorded, but they are based on incomplete and oversimplified information models to effectively represent key archaeological data concepts and capture some vital types of information.

Chapter 3 – Archaeological Methodology

Chapter 3 discusses the methodology used in the investigation of this thesis. The investigation of archaeological database requirements consisted of several concerted methods. The first method included research analysis of specific databases and popular database system software currently in use for archaeological collections. The second method was to investigate archaeological manuals and other archaeological instructional texts regarding proper procedure used for data collection and recordkeeping. Third, project methodology also included interviews with curators, registrars, and other types of archaeological collection staff.

Commercial social science collection information systems are by-and-large designed for richer collection disciplines such as art; and they have been co-opted for archaeology because of the general similarities that exist between the collection types. All types of cultural collections have similar administrative data requirements; e.g. all require as part of good curatorial practice recording receipt of an object and treatment of it while in their collection. All collection disciplines also share the basic requirements for effective database use, such as the ability to efficiently enter, retrieve, and share data about collection objects.

However, databases designed for biological samples or art objects often lack key archaeological concepts such as Context, Feature, or both. Context and Feature directly contribute to understanding the significance of an archaeological object, and these concepts differentiate archaeology from other types of cultural collections.

Chapter 4 – Comparison of Biological and Archaeological Collections

Chapter 4 reviews the conceptual similarities and differences between archaeological and biological collections. Archaeological and biological collection types differ in classification and organization methods, and the importance of locality. The classification method for biological collections is ultimately based on a nomenclatural system derived from the publications of Linnaeus. The Linnaean derived system (i.e. an agreed upon set of nomenclature protocols for applying latin terms for hierarchical arrangements of related organisms) is essentially universal. However, no universal classification method exists for archaeological collections. Biological collections are also organized based on classical divisions of biological classification (e.g. herpetology, ichthyology, mammology). This method of administrative and physical organization of biological specimens creates divisions in their collections which have highly similar data requirements.

Data associated with archaeological collections in contrast are highly variable. One single collection may contain objects that can be separated into literally hundreds of different categories depending on how the collections manager or curator decides to classify them. Finally, archaeological collections often require documentation with much more precise locality and site information. Archaeology data also typically requires more security, with carefully controlled access – more data are needed for biological or art collections.

These requirements, both similar to, and different from, biological collections, were used to inform the design of an information model for archaeology. An information model is a conceptual representation of the way in which classes of information are related to one another within the context of a research domain. The goal of the design was to create a database which would facilitate not only collections management but also research efforts. It was not the intention of this effort to force a single design or management system implementation upon collections but rather to try to best accommodate the way that data are used today, to anticipate how it might be used in the future and to encourage the collaboration of researchers.

The information model proposed here is intended to allow the level of detailed record keeping required of museum collections as stewards of the public interest in regards to both administrative and object-specific information. It is further intended to support research interests by answering basic searches and complex research questions.

<u>Chapter 5 – A Model for Adapting Specify for Use with Archaeological Collections</u>

Chapter 5 explains the adaptations that must be made to the current Specify biological data model for it to be effective for archaeological collection management. The investigation determined that Specify could be adapted to effectively describe and catalog archaeological collections. However, several significant modifications would be necessary for Specify to become an effective archaeological database management environment. These changes would especially include accommodation for key archaeological concepts of: provenience, feature, and context, all of which are not adequately implemented in Specify or any existing information system. However, there is at least one desirable characteristic that the model derived here does not attempt to incorporate--commonality of terminology and categorization. Specify does not attempt to create an archaeological dictionary of terms and categories (authority control) that all archaeologists can agree upon thereby allowing greater sharing of data.

The following thesis begins by reviewing the progression of databases in archaeology, including their use and technological development. Next, I describe the methodology used in the investigation of the thesis, including definitions of words used throughout, methods of investigation, and a comparison of biological collections and archaeological collections. Finally a description of the information model created is described in detail as well as an explanation of the changes made to the current Specify model.

Chapter 6 - Conclusion

Chapter 6 concludes with steps that must be taken before the proposed information model can be made into a fully functioning software program. The chapter discusses the potential advantages offered by the proposed information model as well as its potential weaknesses. A domain information model is only the first step in a multi-step process towards the implementation of actual software.

Chapter 2 – Review

Efforts to create a model for archaeological collection database systems and the corresponding software programs are approximately concurrent with the evolution of computing technology and practice going back to the 1930's. Efforts have waxed and waned but have remained present in one form or another since the advent of the computer. Like other collection disciplines, archaeological information management has experienced cycles of theory, conceptualization, and implementation.

The ongoing development of increasingly sophisticated computerized databases generally tracks technological advances. With every advance in information technology, there is a subsequent revitalization in interest to create more efficient and improved collection management systems. In the cycle, a theory of database design is proposed debated for a period of time and is typically followed by the development of software programs, which serve as prototypes or in some cases mature to become as production systems.

Each generation of database programs is used until it becomes apparent that software is deficient in fundamental ways, at which point a debate of database design concepts and technologies is reinvigorated. There is little push for design during the periods of theoretical debate, though some programs are produced. These periods are quickly followed by fallow periods where the focus of attention is on data integration. At regular intervals during these debates, calls for integration and for the creation of a universal database typically are made to have the records from multiple databases compiled so that only one source exists for information on either the objects from a particular region or from a larger area, such as a national database.

Such a cycle began with one massive advance in technology, the creation of the modern computer in the 1950s. In-depth discussions of the potential usefulness of computers followed from the 1950s into the 60s when museums began to consider computers as useful tool for managing collection databases.

The mid-1960s ushered in a wave of institutionally-based experimentation with computers in archaeology (Gaines 1984, Krakker *et al*. 1999). Several initiatives attempting to address collection data management issues with computer-based systems emerged. One was the University of Oklahoma GIPSY software, designed for ethnographic collections and intended to spark a nationwide push for computer-based systems. However, GIPSY failed to find support outside of Oklahoma and Missouri and shut down soon after it was created (Sarasan 1981).

George Cowgill created one of the largest and most extensive databases for the Teotihuacán Project conducted by Rene Millon from the University of Rochester. Cowgill used this database to analyze the ceramics collections and other objects recovered from systematic surface collections in order to infer significant patterns (Renfrew and Bahn 2004, Cowgill 2008). It is now one of the most valuable tools for research into the site of Teotihuacán and the Teotihuacán culture in general. Although Cowgill's database was groundbreaking, the system he created was not used by other collections.[1]

Despite these projects, it was not until the 1970s that the pursuit of digitization begun in the 1960s was pursued on a large scale (Gaines 1984). This upward shift in momentum of digitizing data into practice was facilitated by technological advances in how databases functioned. Up to that point computer-based databases consisted of flat file data banks. This record keeping system offered little assistance to users in data entering, accessing, and searching. It essentially mimicked in electronic form the storage approach of the card catalog system (Lock 2003). However, a database technology paradigm shift occurred in 1970 when E. Codd proposed the concept of a relational database (Yeung and Hall 2007).

[1] There may have been some exceptions but none that could be found mentioned or referenced in literature.

Codd's relational data model revolutionized computerized databases as a useful means of managing collections. He observed that using a relational database would solve the leading information management issues encountered by museums (Codd 1970). Instead of a special purpose, analysis-focused database, Codd's relational concept based on data tables which could be linked and joined as needed for ad hoc queries offered a more flexible design approach, and could accommodate changes needed to meet the evolving requirements of database users.

U.S. Federal agencies increased interest in the use of relational databases. In the 1970s, Cultural Resource Management programs pushed to create a more streamlined method of doing inventory. They developed computer-based automated inventory systems and a subsequent drive to create archaeological collection database systems gave rise to a series of discussions on what the true capabilities of computer-based databases might be (Gaines 1984).

Following a shift in priorities from collecting and toward museum record keeping and the potential of computers to integrate archaeological collection data, U.S. museums focused on creating a universal, or at least national, archaeological collection database (Sarasan 1981). In order to have such a database and avoid pitfalls that befell the University of Oklahoma GIPSY program, a consensus had to be reached regarding what information was vital to an archaeological database and what archaeological data should be collected and stored (Sarasan 1981). To that end, a group of archaeologists experienced in database management met in 1971 at the University of Arkansas (Chenhall 1981, Gaines 1984), to address relevant issues.

This conference, the first Archaeological Data Bank conference ever held, discussed broader implications and potentialities of computers in archaeology. However, the primary goal of the conference was to discuss what data were collected by archaeologists working in different areas of the world and to produce a list of the data to be stored for their collections. In the end, conference participants could not reach a concensus (Gaines 1984). They concluded that archaeological data were too disparate to be consolidated into one universal short list of definitions and types.

Following that conference, the archaeological community kept the goal of an integrated computer based database system at arm's length. Instead, a series of institutional specific systems were developed, including SELGEM created by the Smithsonian Institution (Sarasan 1981). Within the institutions with computer-based database systems, the systems were little more than a support databases; they were not the primary source of data for the collections or the objects. Databases were essentially a backup system for other types of record keeping (Chenhall 1981). Continued interest in creating an integrated computer based database system was demonstrated by publications like *For Concordance in Archaeological Analysis: Bridging Data Structure Quantitative Technique and Theory* by Christopher Carr (1985).

By the 1980s it was still recommended that computers be used only as a backup for collection management in conjunction with the database record cards. Computer-based databases were essentially a tertiary support after card catalogs and unwritten shared knowledge of artifact history and location (Sarsan 1981). Database users were advised that computers were best used as a cross indices for paper records (Chenhall 1981). The computer records contained only enough data to locate the paper card that contained all of the data on the collection object (Krakker *et al.* 1999).

Nevertheless, during the 1980's computer technology continued to make great commercial advances and with it software and database systems. Computers became more easily accessible and affordable units and were no longer the massive mainframes which required entire floors or rooms to be used. Suddenly, computers could fit on an individual's desk. This one innovation exponentially increased the interest in computers and computer programs (Krakker *et al*. 1999). Databases became increasingly computer based and computer software for general databasing needs was designed. However, as Saransan (1981, 42) stated, "it is apparent that in the haste of some institutions to '*do something*' they have not necessarily done it well".

During the 1980s, Harold Dibble and Shannon McPherron studied data integration needs for archaeological collections. They listed the requirements necessary for data integration, including compatible database structures. They briefly discussed the requirements for a database to be an effective tool for data integration and stated that the database must be flexible enough to accommodate the specific needs of different institutions and different types of materials (Dibble and McPherron 1988). Further, different phases of archaeological projects, *i.e.* excavation, conservation, curation, and analysis, must be integrated. Ideally, each of the phases could enter their data into the same database.

However, during the 1980s relational databases began to be preferred over the flat-file style (Olien 1992). Though the advantages of the relational database system had been touted repeatedly since Codd's initial work in 1970, flat-file systems with no information or logical relationships among them remained the preferred style. The flat-file system was comfortable for many because it reproduced the card catalog and a visual organization of data with which many researchers were familiar. However, by the mid-1980s, the advantages of the relational database were widely recognized and relational systems started gaining commercial success. As a result, computer-based database systems began to reflect the relational data model concepts over flat-file modes of data organization.

Despite the extensive 1980's discussions of the needs of archaeological database requirements, programs created following the debates did not heed the pronouncements. In the early 1990's many database systems were created for businesses but few, beyond institutionally designed systems for in house use, were created for the museum community. Some database development environments for personal computers, including dBASE IV from Ashton-Tate, were marketed for business use but they were adapted by museums for documenting collections (Olien 1992). At that time, museums still commonly used flat-file databases. Analytica created Reflex, one of the leading database programs at the time, a flat-file style popular for the PC (Olien 1992). Another "simple" data processing system which was frequently used by museums was the spreadsheet program Lotus 1-2-3, which was highly popular on the first generation IBM-PC (Darrow 2002).

Museum collection data management software created in the mid- to late-1990s was almost exclusively relational; it incorporated newer highly useful technology which had become vital to the discipline of archaeology such as Geographic Information Systems (GIS) (Arp 2004). In the late 90s, museums became a target market for database software designers. As a result, programmers began to orient database management programs towards the unique needs of museum collections. PastPerfect was developed in 1998 specifically for cultural collection data management (http://www.museumsoftware.com/) and has enjoyed success among museums to this day. Though database environments were specialized for museums, they tended to be general and useful towards all types of cultural and historical museum collections. Museums were eager to employ one program across different collection types and departments, to simplify database support. However, in the quest to create an interdisciplinary database environment, the ability of the software to be highly customized around the specific needs of the departments suffered.

Some archaeologists felt constricted by database software, believing that it did not take into account significant archaeological concepts such as context. Consequently, archaeologists were relegated to use general text fields to enter essential data. “Remarks” and “Notes” fields were filled with long sections of text that were unstructured, uncontrolled and which could not be effectively searched (Leith 2008). Because of that lack of utility, data were often left out of the databases, and could be found only in field notes.

During the late 1990s and 2000s, there was a shift in the focus of articles regarding archaeological collection databases away from database programs. Though interest in database environments was still present and they continued to be innovated, they started to become less important than other interests, such as the potential of XML (a universal computer language for structuring data), GIS, and data integration over the Internet using network protocols.

The goal of creating a universal archaeological database has experienced a recent resurgence of interest. Databases are now seen as a tool, a means of achieving the ultimate goal: integration or aggregation of records from different information data providers into a single database. Scope of large-scale data integration visions have ranged from objects from a region, such as the Aleutians, or a larger area such as a nation. As a result, much of the work done in the late 1990s and 2000s heavily focused on data integration with some mention of database design and implementation, mostly as a side note before discussing integration.

Archaeological journals were created to deal with advances in communication and their effect on archaeology (Snow 2006). Such publications, including Internet Archaeology, give great deference to data integration and therefore computer-based database systems, in particular their potential role in data integration. Several initiatives also began to emerge in the late 1990s and early 2000s with the specific purpose of encouraging and facilitating data integration.

Several government programs have been recently initiated to encourage the integration of the databases for areas directly affecting archaeology such as GIS data (Snow 2006).

Several groups such as ones led by professors Keith Kintigh in 2004 (Kintigh 2006), Dixie West in 2007, and Herbert Maschner in 2011 took up the challenge in the 2000s. Numerous roundtables and workshops were created to discuss the necessity and requirements involved in creating a national database. Ironically, these meetings calling for integration were in large part completely independent attempts. Programs such as the Archaeology Data Service (http://ads.ahds.ac.uk/) and Open Context (http://opencontext.org/) were created to provide an open forum on which datasets can be published. They are intended to provide access to the datasets that usually are only available through individual institutions or their websites.

Another organization, the Society for American Archaeology in 2006 created the Digital Data Interest Group (Kansa February 2007). The DDIG is intended to examine the way in which information is handled and shared with the goal of fostering better communication and access. The DDIG was created through the efforts of Keith Kintigh, Dean Snow, and Eric Kansa. Their part in the creation of the DDIG seems appropriate as they are three of the leading archaeologists writing articles touting data integration.

The principal interest of both the DDIG and the work done by its creators is data integration. Databases have tagged along with the debate as a necessary part of data integration but have not been debated great detail. In fact, databases and data management have fallen back as a 3rd level concern, data acquisition methods have increased in priority as intellectual debate concerns over data management.

Although the debate has shifted away from database design concerns, if the interests of the DDIG are any indication, it is likely to return. The cycle of interest brings the discussion back to data management and databases. Once the discussion of the importance of data integration has become established, the question of how that will happen will have to follow and be brought to the forefront. In the meantime, database models for archaeological collections, such as the one proposed in the current thesis, will continue to be analyzed and argued hoping to address the data management needs of archaeological collections.

Chapter 3 – Archaeological Methodology

In order to consider Specify for archaeological collections, several preliminary intellectual steps must be undertaken. An effective information model for archaeological collections database management systems (DBMS) cannot be created until an understanding of the requirements for information retrieval, representation, and integration is attained. Archaeological databases serve several functions; they are used to record the information collected by archaeologists in the field and in the laboratory, they serve as research resources for future archaeological projects, and they are used to track the condition and disposition of collection objects for administrative and legal purposes.

In order to understand the requirements for an archaeological collections database for all of these uses, I took a three-pronged approach. First, I surveyed data collection and record keeping techniques in archaeology both as recommended and as applied. Second, I interviewed archaeologists and archaeological collections curators regarding how their collections are used, by themselves and by researchers, and the DBMS they use. Third, I examined the currently available DBMS's to understand what is available and what their strengths and weaknesses are. These three dimensions will be used in the next chapter in an examination of the similarities and differences between archaeological and biological collections, for the purpose of determining changes that must be made to the current Specify information model.

Archaeological Data Collection Procedures

For my survey of archaeological data collection procedures, I examined both the procedures as they are recommended and taught, and as they are practiced. I began by reviewing several texts on archaeological data collection and recordkeeping to ascertain the recommended data collecting procedures. A second aspect of the data collection procedure survey was to draw upon actual field experience, both my own and that of another archaeologist, Dixie West. The actual experience demonstrates procedures as they are practiced in two areas of the world, the north Pacific and northern Peru.

As mentioned in the introduction, the audience for this thesis is both archaeologists, and computer programmers who have no or little experience with archaeology. For this reason, I have included extensive detail regarding the archaeological practices and principles described both in the texts and the field experiences. Most archaeologists are familiar with the practices and principles details described in the following section, but they are invaluable in creating an archaeological collections DBMS.

Archaeological Field Manuals

Innumerable texts on archaeological data collection are published. I chose to review a small number of these books including *Archeology: Basic Field Methods* by Stewart (2002), *Archaeologist's Fieldwork Companion* by Kipfer (2007), *Fundamentals of Archaeology* by Sharer and Ashmore (1979), *Excavation* by Carmichael, Lafferty, and Molyneaux (2003) in the Archaeologist's Toolkit Series, the *Handbook of Archaeological Methods* by Maschner and Chippendale (2005), and *How to do Archaeology the Right Way* by Purdy (1996). The texts describe a multi-step process for data collection. The steps include research, survey, excavation, classification, and analysis. Each step involves unique procedures and data collected.

Archaeology: Basic Field Methods (Stewart 2002) is intended to be a textbook for an archaeology course. The book offers what archaeologists would consider a superficial description (non-archaeologists would consider it a detailed description) of many different topics related to fieldwork. However, that is where the book ends. It does not describe what to do with the objects once they have been collected and shows little interest in records other than to say that information should be recorded. This is where the *Archaeologist's Fieldwork Companion*

(Kipfer 2007) comes in. It is focused on offering practical information and sample forms for archaeologists. The focus of the book seems to be on forms more than anything else. *Fundamentals of Archaeology* (Sharer and Ashmore 1979) is also set up like a textbook but it is not focused on excavation. It discusses basic excavation methods but it also examines what to do once the objects are excavated. It is much more vague than *Archaeology: Basic Field Methods*, but it covers a broader span of an object's treatment once discovered.

Excavation (Carmichael, Lafferty, and Molyneaux in the Archaeologist's Toolkit Series 2003) focused on the destructive nature of archaeology. It does not go into depth on any issue but it covers not only excavation, but processes before excavation and the types of analysis that might be done. *Handbook of Archaeological Methods* (Maschner and Chippendale 2005) is a long, two-volume work comprised of articles on excavation and object treatment. The book focuses on analysis and interpretation with excavation and curation only receiving cursory treatments. In fact, curation is only described in vague terms, mostly just a description of what a database is. Although focused very much on regional archaeology, *How to Do Archaeology the Right Way* (Purdy 1996) describes some of the general principles in archaeology and is very concerned with measuring provenience. Somewhat disturbingly, none of the books seem very concerned in curating the objects. A lot of interest is shown in collecting information but not what to do with the information once collected.

Research

The first step in the data collection process is research, which does not concern collection management software so will not be described in detail. Research may consist of examining maps of areas, written accounts of previous exploration, field notes of other archaeologists, interviews, and collections examination (Sharer and Ashmore 1997, Stewart 2002). This step provides the archaeologist with an idea of where potential sites may be located and what finds may be discovered there.

Survey

The second step in the data collecting process is surveying an area of interest identified in research. Surveying often is limited to surface survey but may also include subsurface survey to assess the potential of a site. A surface survey can be done visually or through remote sensing. Remote sensing may include methods such as magnetometry, electric resistivity, electromagnetic conductivity, ground penetrating radar, or satellite remote sensing, to name a few (Kvamme 2005).

Visual surface survey can be run formally or informally. Informal surface surveys consist of a person or persons walking an area and collecting objects without marking their locations. This method produces only general provenience information for the objects. Formal surface survey consists of a more methodical collecting method and a more precise provenience being recorded for objects. Formal surface survey consists of marking every object found. The area can then be either photographed, mapped, or both. For extremely precise measurements, a total station may be used to mark the location of each object (Sharer and Ashmore 1979, Carmichael, Lafferty, and Molyneaux 2003).

Survey may also consist of subsurface survey. Test pits are excavated similarly to normal excavations but on a much smaller scale. They may consist of a shovel test in areas with a shallower archaeological record or a shaft several meters deep in areas with a deeper archaeological record (Sharer and Ashmore 1997, Stewart 2002). Objects discovered in a subsurface survey are recorded similarly to objects found in normal excavation and can include as much detail as objects collected in an excavation (Stewart 2002). The result of surveying is that it either helps confirm or contradict the archaeologist's interest in an area as a potential site.

Excavation

Once a promising area has been identified, the excavation begins. There is no single excavation method or process that works for all sites (Carmichael, Lafferty, and Molyneaux 2003, Glassow 2005), but archaeologists, regardless of region, agree on the necessity to observe and document as much detail about the excavation process as possible. Because archaeology is a destructive act, there is no second chance to document information about the conditions of the matrix being excavated. Therefore, careful attention must be paid to documenting the information for posterity (not to mention his or her own research). An archaeologist cannot go back to the site and re-observe an object's context. The excavation techniques used to document matrix conditions and object context may vary from site to site, but there are some methods that are shared my many archaeologists, such as use of a datum and grid and excavating in layers.

The first step for many excavations is to set up a means of coordinate control. This is often accomplished by establishing a site datum (Carmichael, Lafferty, and Molyneaux 2003, Glassow 2005, Purdy 1996, Sharer and Ashmore 1979; Stewart 2002). A datum, a standard reference point for the site, cannot be disturbed or moved. The datum has known coordinates and elevation measurements from which the coordinates and elevations of excavated objects are ultimately derived. This point is also used to establish the corners of excavation units (Purdy 1996). Subsidiary datum points may also be established within a site, the coordinates and elevation measurements for which are derived from the site datum point (Glassow 2005).

Before breaking ground, archaeologists may establish a perimeter with known coordinates or a grid within the perimeter. The use of a grid enabling precise measurement has been identified by at some archaeologists as the defining point in modern archaeology (Carmichael, Lafferty, and Molyneaux 2003). Setting up a grid consists of establishing grid lines over an excavation unit, usually in a quadrilateral form (square or rectangle). The grid may only be put up when measurements are taken and objects and features are mapped, then taken down immediately (Carmichael, Lafferty, and Molyneaux 2003, Stewart 2002).

The corners of the excavation and the grid are marked with stakes or nails, something that will not move during the period of excavation. The coordinates and height of at least one corner of the grid is established based off of the datum. According to Purdy, the southwest corner of the grid should be measured and used as a kind of subsidiary datum, and should be marked on each object bag (1996).

After the boundaries of the unit are established, the excavation can begin. Controlled excavation is dug in layers. The layers can be dug in either natural or arbitrary layers or a combination of the two (Carmichael, Lafferty, and Molyneaux 2003). Excavating by arbitrary layers consists of digging by a set depth for each layer, such as 10cm. The layers can create a level floor or a floor that follows the natural topography of the surface. Measurements must be taken for each layer, even when excavated using arbitrary layers.

Excavating by natural layers consists of following the naturally occurring layers, the beginning and ending of which is marked by changes in the matrix (Stewart 2002). Natural layers vary in depth and a single layer may vary in depth from one end of the unit to another. Multiple elevation points may be taken for each layer in order to keep track of context. Matrix characteristics are particularly significant for natural layers. Changes in the color and texture of the soil not only indicate the beginning and ending of layers, it indicates the presence and boundaries of features. Excavating by a combination of the two methods follows the natural layers in the unit, but if the layer is too deep, arbitrary layers are used.

Object Recovery

When an object is discovered, it is exposed and cleaned. Before it is removed, the object should be photographed *in situ* and mapped (Carmichael, Lafferty, and Molyneaux 2003, Purdy 1996, Sharer and Ashmore 1979, Stewart 2002). An object is mapped by measuring its location using the grid and taking the object elevation. Precise measurements are taken for each object giving them a point provenience. The measurements are used in conjunction to identify precise coordinates for each object (Carmichael, Lafferty, and Molyneaux 2003). If it is sizeable, several points are taken to properly depict it in the map (Carmichael, Lafferty, and Molyneaux 2003, Purdy 1996, Sharer and Ashmore 1979, Stewart 2002).

The grid used for excavation correlates to lines or points on the map. The map is marked with known coordinates or associated with points whose coordinates are known. An object may be mapped on several maps. The object number is noted on each map so the associations and context can be preserved (Carmichael, Lafferty, and Molyneaux 2003, Purdy 1996, Sharer and Ashmore 1979, Stewart 2002). The object elevation may be marked on the map, on the object record, or both. If the object is at an angle, a top elevation and a bottom elevation may be taken (Sharer and Ashmore 1979, Stewart 2002).

If a collection of objects is found, such as stone flakes or debitage, the objects may be collected as a lot (Purdy 1996, Sharer and Ashmore 1979). Rather than being mapped individually, they are mapped as a group. Several measurements around the periphery of the lot may be taken to define the area from which the objects were collected. If there are objects in the lot that are deemed to be significant, then the significant objects will be treated individually and the rest of the objects will be treated as a lot (Sharer and Ashmore 1979). The significant objects will be mapped, collected, and documented individually. The same approach is taken on objects collected from screening and flotation. Since the precise provenience cannot be collected, a general provenience is given to all of the objects (Purdy 1996, Sharer and Ashmore 1979). If a significant object is discovered in the screened or floated material, then it will be collected and documented individually.

After the objects are collected, they are placed in a container, usually a bag or a box, unless they are too large for a container. Objects are placed in individual containers, with the exception of lots which are placed in the same container (Purdy 1996; Sharer and Ashmore 1979, Stewart 2002). Objects of different types (*e.g.* bone, ceramic, metal, etc.) may be placed in the same container if lots are not sorted in the field (Sharer and Ashmore 1979). Objects from lots deemed to be significant are placed in separate containers from the rest of the lot (Sharer and Ashmore 1979, Stewart 2002).

An object number or a catalog number is written on the container in which the object is placed (Kipfer 2007, Sharer and Ashmore 1979, Stewart 2002). If the object is too large for a container, the information is often written on a tag and somehow attached to the object in a way that will not damage the object. The object number can be an arbitrary number or consist of significant numbers (Kipfer 2007, Stewart 2002). The format of the object number is not as significant as the ability to associate it to its provenience and context information. The object's number or a record with some provenience information may also be placed in the container or attached to the container. The object number or catalog number is the number recorded in the archaeologist's field book as a backup for the record attached (Sharer and Ashmore 1979).

Description and Documentation

The description of the records used to document object context varies greatly from text to text. The commonality of all of the descriptions is that the record is intended to document as much detail as possible. Documentation collected in the record may include details about object characteristics, but it also will include details about the context from which the object was collected and details about its collection. Some of the information recommended for documentation includes the object number, excavation method used to discover the object, collector's names, date it was collected, the provenience when collected, the site, unit, layer, level, quadrant in which the object was discovered, and a general description of the object (Purdy 1996, Sharer and Ashmore 1979).

The records can document more than just object information. Carmichael, Lafferty, and Molyneaux (2003) recommend using layer records to document information on each layer, level records if levels are used, and feature records to record information on features. Sharer and Ashmore (1979) recommend the use of unique forms for the different recovery method, *i.e.,* surface survey, shovel test, test pit, excavation, etc.

A standardized form may be used for most object types, but burials are an exception. They may have unique records with a unique numbering system identifying the gravesite from which they were collected (Purdy 1996, Sharer and Ashmore 1979, Stewart 2002). Human bodies are given particular deference and attention. Stewart (2002) recommends documenting as much about the body as possible along with the type of burial *i.e.,* individual, bundle, cremation, etc. Details collected regarding the body include the position and orientation of the body as well as preservation, completeness, and the objects that were associated with the skeleton (Kipfer 2007, Sharer and Ashmore 1979, Stewart 2002).

Features

Features are another special case in archaeology. They are treated with particular interest because they are evidence of human action, but by definition cannot be collected. Many types of features such as bone pits, roads, or post holes are destroyed when excavated. They cannot be taken back to the lab and studied later. For example, you can bring the dirt from a post hole back to the lab but it will never again be the post hole. Features must be documented as they are being excavated. Features can be excavated in whole or in part depending on the type of feature and the intent and interest of the archaeologist excavating it (Stewart 2002).

Features are assigned a feature number. The feature number may be a continuous number from the beginning of excavation at a site, from the beginning of excavation at a unit within a site, or it may restart for each layer (Kipfer 2007, Sharer and Ashmore 1997, Stewart 2002). If a feature is found within another feature, it is given a unique number from the feature in which it is found. Information that should be recorded for a feature according to some archaeological manuals include; the feature number, drawing numbers depicting the feature, how much was excavated and why, the state of preservation and deterioration, an interpretation of the feature, objects found in the feature, and a general description of the feature's attributes including size, shape, material (Kipfer 2007, Stewart 2002).

Objects collected from features, including samples, may have a different significance from other objects collected in the same layer, because they are in the feature. Objects collected from features may be placed in a separate container from other objects collected in the layer, and the feature objects may have a feature number added to their object number (Kipfer 2007, Sharer and Ashmore 1997, Stewart 2002).

Laboratory Analysis

Once the objects are documented analysis begins. Analysis begins with classification (Sharer and Ashmore 1979). Classification is the process of identifying objects by their attributes including both physical attributes and inferred attributes such as function. Objects can be classified by stylistic characteristics, its form, or the technology used to make it (Sharer and Ashmore 1979). The attributes used for classification depends on the goal of the archaeologist. Different research projects may be interested in different attributes of the same object. Inferred classifications of objects can change over time or from archaeologist to archaeologist (Kipfer 2007).

Some attributes can only be determined through tests run on the object. There are innumerable analyses that have been and can be done on objects. There are too many analyses to describe or even list, though most tests are run for purposes of dating, sourcing, and chemical analysis (Kipfer 2007, Sharer and Ashmore 1979). The significant aspect of the tests from a collection management standpoint is that the results must be associated with the object. Unlike excavation, analyses can usually be repeatedly run. However, running an analysis causes some damage to an object or loss of an object; therefore, if an object is analyzed, the analysis is often recorded for the object so another damaging analysis does not have to be conducted to establish the same information (Sharer and Ashmore 1979, Stewart 2002).

Summary of Procedures

Based on the recommended procedures and practices several key points can be gleaned. First of all, while there are common guiding principles, archaeological approaches and styles vary greatly. What information an archaeologist collects and how he or she collected it can be very different. Second, once an area is excavated, no further excavation data can be collected from that area. If data on context and matrix conditions are not recorded, they are lost forever. Third, it is recommended to archaeologists that they record information in as much detail as possible, which includes recording information in as great a variety of mediums as possible. Fourth, precise measurements should be taken, including provenience measurements. Fifth, context is extremely important. Archaeologists should take great care in documenting context in as detailed a manner as possible. Sixth, features are a unique case and are treated differently from other objects and contexts. Seventh, objects may be cataloged as a single item or in lots, depending on the methodology used and the nature of the objects themselves. Eighth, objects may be classified in a multitude of ways depending on the archaeologist's needs. Finally, there is a limitless number of analyses that can be run on objects, many of which may not yet exist.

Field Experience

My experience with archaeological procedures applied in the field relied heavily on my two year experience digging in northern Peru with Dr. Luis Jaime Castillo Butters, Roxana Barrazueta Pino, and Julio Rucabado-Yong. Castillo Butters has been the Director of excavations at the site of San Jose de Moro in northern Peru since 1991. Julio Rucabado-Yong and Roxana Barrazueta Pino are both former students of Castillo's who both continue to assist him with the San Jose de Moro (SJM) excavations. Rucobado-Yong is a site supervisor for SJM as well as a director of surveys of nearby sites. Barrazueta Pino supervises an area within SJM and is in charge of data recording for the site.

San Jose de Moro

San Jose de Moro is the site of a Moche cemetery and ceremonial center. The site has produced highly complex and intricate burial and ceremonial contexts. The site maintains several concurrent excavations in different areas of the site every summer. Interpretation is heavily context dependent for the site; thus the precise location of each object is very important (Castillo 2008). A dependable means of documenting detailed records is necessary.

These researchers have created a reliable system of recording that they have used for over a decade. The area supervisor assigns consecutive numbers to the natural layers as they are excavated until a sterile level is reached. A picture is taken and a plan may be drawn of a layer if something of interest is found in it. If the item of interest is an object or many objects, the object and the object's context are photographed more closely, drawn in detail, and removed.

A list is kept of object numbers assigned to objects as they are removed. The object number reflects the site code, the area of the site in which the object was discovered, the feature number (if the object is found in one), and the object code. The object code consists of an object material code, i.e. OA for animal bone (oseo animal) followed by a consecutive number. Object materials are separated into their different object types including human bone, animal bone, ceramic, metal, spindle whorl, stone, beads, organic, shell, wood, textile, and soil sample. Thus, the 8th llama bone found in feature3 of layer 15 of area 38 of San Jose de Moro would have the object number SJM-A38-C15-R03-OA08.

A different record form and object code are used for objects recovered from burial contexts. The record form used for burials is very similar to the form used for other objects with the exception of the record title, Ficha de Tumba and the addition of the tomb number field. Similarly, the object code is almost identical with the exception of the addition of tomb number into the code. The resulting code is comprised of the site code, the area of the site in which the tomb resides, the tomb number, the feature number (if the object is found in one), and the object code. The numbering system for feature and object code is separated from other numbers for that layer. The numbers restart for each tomb.

Following collection, the objects are placed in a bag or a box, depending on the size of the object. The object code is written in marker on the bag or box. The object code is written on a record form which has two parts, the record itself and a detachable part which is tied to the container in which the object has been placed. The record includes the object number, the date excavated, the date recorded, the excavator, the recorder, material, and a description of the object.

Drawings of objects and layers are oriented using a grid system. The corners of the grid have identified coordinates. Because the coordinates are known, Castillo and his assistants are able to scan the drawings into CorelDraw (a PC software program for graphics from Corel Corporation) and create a digital map of all of the objects excavated from the entire site in relation to one another. Each drawing code consists of the site number, the area number, and a drawing number. The drawing number is a consecutive number system based on how many drawings have been produced by that area of the site that season. For example, the 8^{th} drawing from area 38 in the 2007 field season would have the number SJM-38-008. Each drawing must have the direction of north, the drawing code, the drawing date, the name of the drawer, and the names of the person or people taking the measurements. Once the objects depicted in the drawing have been assigned object numbers, the numbers are written in the drawing of the object, in order to identify the object's position and context.

At the end of the day, the object records are compiled in binders according to their area, material type, and then object number. The object information is then cataloged in a Microsoft Excel spreadsheet for each area of the site. A list of drawings with a description of each is cataloged in a separate Microsoft Excel spreadsheet.

In addition to being a site supervisor for SJM, Rucabado-Yong also ran survey trips to other sites in northern Peru such as Huaca Salida and Huaca Rajada (Pacanga Vieja). The surveys consisted of digging one or two test pits at each site. The test pits were dug by natural layers until sterile soil was reached. A handheld GPS was used to find the coordinates of one of the corners of the test pit from which the coordinates of objects could be plotted. The documentation system for the survey objects was identical to the documentation system for SJM.

Rucabado-Yong stressed the importance of a field notebook in archaeology for both traditional excavations and surveys. The object records, drawings, and photographs document many significant aspects of an archaeological dig, but the field notebook records much more information. In SJM, the archaeologists' field notebooks served as a place to note soil conditions of each layer, notes on the potential relationship between layers, theories about the objects or levels, or preliminary observations of objects, features, layers, and surrounding land.

Aleutian Islands

The second case study is West's excavation in the Aleutian Islands, described briefly in the introduction. As mentioned in the introduction, West's team is highly diversified. The researchers on her team focus on a variety of aspects of the archaeological sites including the animal remains, the geologic characteristics, and the ecological remains, in addition to the cultural remains.

West's excavation technique is similar to Castillo's but the labeling process is distinct and includes the practice of water screening, which adds a new element and challenge to recordkeeping. West's team is less reliant upon technology in the field than Castillo's team. They work on remote islands in the Aleutians and live with little to no contact with the outside world and no electricity until the end of the field season.

Only one excavation unit is open at a time. Within the unit, each archaeologist is in charge of excavating a portion of that unit. If an object is deemed special, it is cleaned up then drawn in situ. The drawing is done in a notebook specially allocated for drawings. The object is then piece plotted in a level drawing. The catalog bag number is written on the drawing to identify the object depicted.

With several archaeologists all working at the same time, confusion can easily erupt. To reduce confusion, all findings are recorded in a single Official Notebook. A list of object numbers is maintained in the notebook as well as written on the bag in which the object is placed. The object number reflects the site code, the unit number, layer, level, catalog number (object number), object material, date excavated, and collector name. If the object was collected from a feature, the object number will also include the feature number of that layer after the layer number.

The site code is assigned by the Alaska Office of History and Archaeology (OHA). Archaeologists in Alaska can contact the OHA before they go into the field for a list of pre-assigned site numbers. The archaeologist can then assign sites with a number immediately. If the archaeologist does not get a site number ahead of time or runs out of pre-assigned numbers, the site will get a temporary field number, which will be replaced when the OHA can be contacted (Corbett 2009). In cases in which a temporary field number was used, a record will need to be kept of both the temporary field number as well as the official designation number for the site.

Unlike SJM, which restarts the feature numbers at each layer, West keeps a continuous feature count for the site. Features are assigned a number according to the number of features which were found before them in the site as a whole regardless of pit, level, layer, or field season. Feature 110 is the 110th feature found at ADK-011 since excavation first started.

West calculates the coordinates and elevation of a single point or datum for the site. All of the measurements for the site are calculated using this one point. Layers are dug in a combination of natural and arbitrary layers. Archaeologists dig down until they find a change in the soil matrix or 10 cm, whichever comes first. If they did not reach a natural layer break within 10 cm, they level the pit at the 10 cm depth and label the excavated portion Layer X, level X, incrementally increasing the numbers as they dig down. When a change in matrix is reached, the pit is leveled to that level. Excavation beneath the layer is given the next available number. Level numbers restart with each layer.

Drawings are not assigned unique numbers. Instead, the Site, Unit, Layer, Level, Date, and depth below the Datum are recorded on the corner of each drawing. Site drawings (plans of the site) are drawn in a separate notebook from the Official Notebook. Features are drawn in on the site plan. The codes for the objects found in particular features are written on the bottom of the plan. The provenience for the object is then marked on the plan. The provenience is not written on the object bag or the field notebook so the only way to know the exact discovery point of the object is by referring to the plan. This practice makes the site plan even more significant because there is absolutely no way to reconstruct context without the site plan.

Summary of Procedures

The methodologies utilized by Castillo's team and West demonstrate several interesting similarities and differences. Important similarities between both methods include the utilization of several mediums to record information about an object and its context. Both archaeologists use the grid system and record provenience data for key issues. An object number is assigned to each object and the provenience information is documented on drawings, not attached to the object. The object number is the link to the provenience information.

The differences must also be considered. Although both excavate in levels and layers numbered consecutively, West excavates in only one area at a time whereas Castillo has several excavation areas open at the same time. Therefore, West has only one number sequence for excavation layers but Castillo has several sequences in effect at the same time. West assigns numbers consecutively to features from the beginning of excavation whereas Castillo numbers features consecutively from the beginning of each layer. Consequently, several features in Castillo's excavation may have the same feature number, but a different layer association. The approaches of these two investigations highlight the potential variation in data sets and data collection methodology.

Interviews

The second step in understanding database use consisted of interviews of curators, registrars, and other employees of museums with archaeological collections of different sizes regarding their current DBMS, how they document their collections, problems they have with their current database system, and how improved database creation and management could better facilitate their needs. The people interviewed were employees or volunteers of Museum of the North, the Field Museum, the Sam Noble Oklahoma Museum of Natural History, the University of Kansas Archeological Research Center, the Kansas State Historical Society, and the National Air and Space Museum Support Center.

Museum of the North, Fairbanks, Alaska

I interviewed James Whitney, Josh Wisniewski, Kerrie-Ann Shannon, and Dawn Planas of the Museum of the North in Fairbanks Alaska. James Whitney is the collection manager for archaeological collections at the museum and Wisniewski, Shannon, and Planas are assistants in the department. The museum's Archaeology Department currently uses 4D (4D SAS) to manage its collection information. They chose the 4D system because the program is highly user-configurable and it could be used with the Apple Mac platform. The system also allowed them to use a sub-accession number system, an option not offered by any other system they could find. A sub-accession number system is utilized for objects that were not accessioned at the same time but were collected at the same site. The sub-object number would have a core object number followed by a consecutive number for objects from a particular site. Thus, objects collected from ADK-11 during the seventh field season of the site may have the number 1001.07. Objects from the same site during the next field season would have the accession number 1001.08.

Despite these advantages, they found significant issues with the system. First, they were not able to attach photographs of objects. This was particularly frustrating for conservation treatments. The procedure for the department is to take a picture of an object before a conservation treatment and take another after the treatment. Since the DBMS does not allow them to attach pictures, they had to maintain separate locations for object related pictures. Second, there was no searchable field for excavation technique. They noted that the method an archaeologist used to obtain the object tells a lot about the reliability of the object's related data. Third, there is no means of documenting who has edited the object information. They believed it was important track who altered object data in order to judge data reliability. Fourth, there was no way to track changes in object definition. Occasionally an object may be erroneously defined, for example an object defined as a statue is later discovered to be a mask. Their system only allowed them to maintain a current definition for an object. They wanted to be able to record past definitions as part of the object's history.

Field Museum, Chicago, Illinois

I interviewed the Registrar for the Archaeology Department at the Field Museum, Misty Tilson. She reported that records go through many different hands and different departments. Each department deals with a different aspect of the record. For example, the record begins with the Registrar who documents the accession information for the object before it is sent to the curator who will be responsible for the object. If there are permits involved, the legal department will document the pertinent information involved in obtaining the object. The conservation department may need to run tests or treat the object, and any such interaction will be documented in the record. The department responsible for the object enters general information about the object such as measurements, material, culture of origin, inscriptions (if any exist) and the collection of the object (i.e. excavator, project, provenience, etc.).

The Field Museum organizes its collection using sub-object numbers. Sub-object numbers as the Field Museum uses them relates to items related to each other by use and provenience. These objects will have the same core object number followed by a sub-object number. For example, if a chess set comprised of a chess board and 8 pieces is accessioned, the numbering system for the items would be a core number, such as 001, followed by a sub-object number. Thus, the board would have the object number 001.1, and the chess pieces would be assigned individual numbers consecutively from 001.2 through 001.9. However, Tilson admitted that there is no way to record how many sub-objects exist for a core number just looking at one of the records, which she believed would be a valuable asset (Tilson 2008).

The Field Museum documents not only information about the object but also the exhibits in which objects from the museum were involved. The exhibit information is indicated not only on the object record but a separate record for the exhibit itself is kept in a department responsible for exhibitions. The information on the exhibit includes whether the exhibit is temporary or permanent, the number of pieces in the exhibit, shipping arrangements, dates for the exhibit, insurance information, approval information, and a condition report for each of the objects before and after the exhibit.

The Field Museum originally used FileMaker Pro (FileMaker, Inc.) but decided in late 2007 to switch DBMS's because FileMaker Pro lacked several significant requirements. Three requirements in particular persuaded them to change database management programs. First, Filemaker Pro cannot be shared by multiple departments. Each department had to use a different database management program, one created for the department type (*i.e.,* paleontology, biology, archaeology, etc.). Second, Filemaker Pro did not allow for adequate attachment ability. The Field Museum has several different records associated with an object, many that could not be attached to the electronic record. For example, pictures of the object could not be attached. Third, the report options did not allow for the reports as the museum needed.

Sam Noble Oklahoma Museum of Natural History, Normal Oklahoma

I interviewed Liz Leith, Collection Manager for the Archaeology Department at the Sam Noble Oklahoma Museum of Natural History (SNOMNH). SNOMNH currently uses Specify for its biological collections; however, Specify lacks several significant concepts necessary for an archaeological collection. Leith searched for a different collection management system to use for the archaeological collection but could not find any that served her purpose. Instead, she decided to create her own DBMS for their collection.

The objects in the department are separated into three different collections; the main collection, casts of objects, and a teaching collection. The main collection assigns catalog number consisting of site number, state, and catalog number. In cases where the origin of an object is unknown, the number 0 is assigned in place of site number and state. All of the objects from the same layer share the same catalog number, meaning that a single catalog number may describe many objects.

The different collections are separated not only physically but in the database as well. The cast and teaching collections are given an identifying code at the beginning of their catalog number. The catalog numbers are different for the different collections. The teaching and cast collections consist of collection type, material, and catalog number. The catalog number is created by a consecutively assigned location number and a consecutively assigned number from that location. Thus, a teaching collection lithic that was the first to be excavated from location 56 would have the code TC-LT-56.1.

SNOMNH serves as a repository for several different organizations. Because of this, they need to be able to keep ownership records attached to the object record for objects they are storing. Leith commented that the museum does not keep physical copies of site records. The only copy of the records is an electronic record. Leith noted that attaching site data to an object record would be advantageous for searching object records. The accession number is comprised of the year of accession followed by a consecutive number. Each new object or group of object receives a unique number; there is no relationship reflected in accession number if the objects were collected from the same site (as the Museum of the North does). She also reported the need to track changes in storage location. Objects may be moved temporarily for a number of reasons such as research, when an object may be placed in a separate room for days or weeks. Unless a means of tracking location changes, people may not know where to find a moved object.

University of Kansas Archaeological Research Center, Lawrence, Kansas

I interviewed Mary Adair, curator for the Archaeological Research Center (ARC) at the University of Kansas. Adair reported that collections are curated according to accession, site, material type, and year of investigation. The accession number for objects entered into the database consists of the site number where it was discovered, the year it was entered into the collection, and a consecutive number.

There is no consistent cataloging system. Adair defers to the system used by the field director where it was excavated. However, the site number must accompany the catalog number. This is often followed by a sequential number from a total station mapping instrument.

In order to ensure the ability to quickly locate any given object, the object's location is included in the object's paper file. The location description includes the building, cabinet number or shelving unit, and box number/ drawer number/ shelf number. Thus an object in Spooner Hall, in the 24th cabinet, on the 3rd drawer, would be in S24-3.

If an analysis is done on an object such as radiocarbon dating or composition analysis for determining the source of material for the object, a copy of the report from the analysis run is collected by ARC. If the report is sent in paper format, the report is added to the collection's or site's paper record. If the report is sent in an electronic format, the report is added to the electronic record and a paper copy is made and added to the paper record. Because such tests are destructive, careful record is kept of all aspects of the analysis. The record includes the name of who approved the analysis, when it was approved, the method being used for the test, and who requested the test be run.

Adair differentiates between a collection database management system and an artifact database. The DBMS organizes the collection. The artifact database is a spreadsheet of artifact information. It can be added to as needed as more data become available on the object. She used to use FileMaker Pro for her DBMS and Excel for the artifact database. Excel is used because it is a quick and easy tool for recording data. It is uncomplicated and not a lot of mouse clicks or time is required to enter all of the known data for one object and move on to the next object. It is also a user-friendly system with which many people have experience or quickly understand.

FileMaker Pro was selected because Adair began using the program in the early 90's when there were not many collections DBMS's available. FileMaker Pro offered a relational database and good customer support, which was hard to find. Adair switched to using Excel instead of FileMaker Pro because of the high cost of a software license. She said it was cost prohibitive. She had to pay for upgrades for each computer using the program. Now, she uses Excel. She reports that the benefits of converting her data to a relational DBMS do not outweigh the consequences of having to transfer the data into a new system. Since she has a smaller collection with a small number of data to be entered for each object, a complex system seems to her to be unnecessary and inefficient.

Kansas State Historical Society, Topeka, Kansas

I interviewed Chris Garst of the Kansas State Historical Society. The Historical Society categorizes and organizes its collection similarly to the University of Kansas Archaeological Research Center except that they consider function in addition to geography. After the geographic and accession categories used by Adair, objects are classified and sorted according to the function the object is believed to have been designed and used for (regardless of what it was actually used for which may be difficult or impossible to tell). Thus, knife would be classified as tableware even if it was actually used as a screwdriver.

Garst stressed the importance of unanimity in object descriptions and classifications. The Historical Society depends greatly on the help of volunteers for recording object information on paper as well as electronically. There might be any number of volunteers at any given time recording information and volunteer turnover can be high. Each volunteer has to be trained on how to record information and the experience with or knowledge of historical objects may vary from person to person.

For this reason, the Historical Society created its own data dictionary with a list of terms and categories for the objects. The data dictionary is placed near the computers and work table so volunteers and anyone else who needs to use it can refer to it at any time. Garst reported that the ability to have a common reference for everyone to use was invaluable because everyone knows what everyone else is talking about and confusion is reduced.

National Air and Space Museum Support Center, Suitland, Maryland

I interviewed John Miller, a contractor for the National Air and Space Museum Support Center (NASM), a branch of the Smithsonian. He worked in the collections processing unit of NASM, digitizing collections, creating condition reports, and reorganizing collection storage. While he did not create records for objects in the collection, he was responsible for editing and updating the preexisting records, including attaching photographs of the objects, augmenting object descriptions, entering maker information where it was missing, etc. NASM uses The Museum System (TMS) by Gallery Systems for their database management.

There were two main issues he and his co-workers had with the program, its inability to track past storage locations and the inability to record conservation procedures. The main issues Miller and his co-workers faced with TMS was its inability to track storage locations. TMS only allows for the current location to be recorded. If the object is moved temporarily, the information for the object must either be changed, removing the original location in the database, or the location must be left as it is, giving no indication of its current location to others who may be interested in it. There is no means of recording past positions unless the information is placed in a Notes or Remarks field. He also reported that there was inadequate means for recording conservation procedures. The only means of documenting conservation procedures was to describe the procedures in a general Notes field, which means procedures could not be easily searched for and descriptions of the procedures had to be abbreviated to fit into the field.

Summary of Procedures

Several key points were voiced by the museum employees on their DBMS's, which must be taken into account in designing such a system. First, sub-accession/sub-object numbers are used by several of the institutions interviewed. Such numbers are an exception to the numbers followed by the rest of the collection. Second, it is very important to be able to attach multimedia attachments to records. Third, they wanted to be able to record changes to the object record to be able to document accountability, track movement, and document past interpretations of the object. Fourth, the objects were organized in storage according to geography and collection. Fifth, concern for maintaining control of the data entered and the need for a data dictionary. Sixth, data need to be easy to search, such as analysis results and conservation efforts.

Current Database Management Systems

The third step in understanding uses of archaeological collections included research into popular database systems currently being used for archaeological collections. I briefly tried out several of programs that are currently popular for archaeological collections including; Microsoft Office Access 2003 (Microsoft Corporation), Open Context (Alexandria Archive Institute), tDAR (Digital Antiquity), Re:discovery (Rediscovery Software Inc.), and PastPerfect Museum

Software 4 (PastPerfect, Inc.). Because Re:discovery does not offer evaluation copies of its software, I based my evaluation on the company's presentation of software system benefits and form examples.

Microsoft Office Access 2003

Microsoft Office Access is Microsoft's version of a relational database. It has the advantage that many people use Microsoft for their computers and Access is one of the programs that come in the Microsoft package. Therefore, many people will have access to it. There is a charge for Microsoft Office, but most people paid for the package in order to get the other programs in the package, such as Word, Excel, and PowerPoint, rather than for Access. Access is a side benefit.

Access has the further advantage in that it offers its users a number of options so they can customize the program for their needs. For example, forms can be created using the Design View which creates forms from scratch, the Form Wizard which allows users to create personalized forms using pre-set elements, or one of five AutoForm options which offers pre-created forms. The five options are Columnar, Tabular, Datasheet, PivotTable, or PivotChart.

The highly customizable nature of the program allows users to add whatever fields, tables, and relationships that they want for their database. They have the option of creating a form for context if they want one. They can then relate the form to any other table to which they think it should be related. Users can make the database as complex or simple as they want it to be. If they do not have much data to be entered, they can create a simple database. If they have more data they want to be entered, they can make the database extremely complex.

However, the process of trying to use Access and create the tables and relationships is complicated and intimidating. Before users even begin using Access they need to spend hours or days setting it up (for some people, possibly weeks or months). Users cannot just start with a default database and edit from there. They must first choose a database type, table type, and choose the relations each table uses. If users want a simple database, they still have to deal with a confusing setup process. In fact, workshops are held to teach users simple database tasks such as creating forms, creating searches, creating records, editing records, and creating table relationships (such as the workshop held at the University of Kansas in 2008). However, once this process is completed for a form, users do not have to repeat it for that form. The process does have to be repeated for all forms but only once per form, if the result is what the user wants.

Open Context

Open Context is a software system released in 2006 (opencontext.org). It was developed as a result of a collaborative effort between the Alexandria Archive Institute (AAI) and the OCHRE project at the University of Chicago. Open Context was created for "the electronic publication of primary field research from archaeology and related disciplines" (Open Context). The intent of the program is to allow archaeologists to easily share their data immediately and to test new tools and data sharing techniques. Open Context is particularly noteworthy because it was the first online archaeological system to fully utilize Creative Commons copyright licensing (Kansa 2007).

Open Context offers online access to its system with actual raw data from excavations in many locations across the globe. It allows for different objects types and forms to be recorded. It uses non-specialized fields to describe object data, enabling more objects to be included. However, it also means that the information is less specific for objects. It allows for images to be attached in several locations including object, site, and lot.

Open Context includes context and does so with more specificity than most systems on the market. However, context is not very descriptive. Context is limited to Country, Site, and Lot. It does not offer a means of recording object relationships to one another, nor does it offer a means of documenting precise provenience information. It can record images of objects and the site, providing an idea of context if provenience is recorded.

The Digital Archaeological Record

Similar to Open Context, the Digital Archaeological Record (www.tdar.org) allows for the publication of archaeological finds. The Digital Archaeological Record (tDAR) is overseen by Digital Antiquity and is run out of Arizona State University (tDAR 2012). tDAR was created as part of Digital Archaeology's goal to "improve substantially the ease of accessing and using archaeological information" and "to provide for the long-term preservation of the irreplaceable records of archaeological investigations" (Digital Antiquity 2012).

tDAR is a free, collection based repository system for archaeological data. It is open to anyone with access to the Internet to upload or search archaeological data. Users can upload descriptions of their collections, related documents in a variety of formats, images in a variety of formats, and spreadsheets with data from the projects. It is compatible with metadata standards such as DublinCore and MODS (tDAR 2012). However, because it allows the user to upload a spreadsheet they created themselves, it does not restrict what data can be documented. Also, because each collection is independent of each other, each project can document different data as determined by the archaeologists on that project.

Users can search collections and view any of the associated data that has been uploaded. Users can also download data that is of interest to them. This allows users the potential to quickly find hundreds or thousands of records fitting user designated parameters in one search.

However, the user may have to search through many spreadsheets to find the data. It is not presented in a readily comparable or compiled format.

tDAR offers a readily accessible way to share data but it is not a relational database system or collection management software. It is not meant to be used to document administrative data such as location in collection, accession information, etc. It also does not allow the user to take advantage of a relational data model, which will be discussed later in the thesis.

Re:discovery

Re:discovery (www.rediscov.com) began in 1988 under a different name with the specific intention of managing the Thomas Jefferson Foundation's collection. Unlike Open Context, Re:discovery is actually collections management software. It was created with the intention of facilitating management and interpretation of objects, particularly historic objects (Richardson 2009).

Re:discovery offers many advantages for research goals and for general collection management. It allows electronic files to be linked to object files. Articles related to an object can be linked to the object along with images of the object, or any other significant images. It also allows the image to be shown when the record is opened. Instead of having to find the image attachment and open the link, the image is treated like other data fields which appear automatically.

It also incorporates a data dictionary for the objects in the collection so continuity can be encouraged. The data dictionary is editable, which enables the user to add terms if the dictionary is found to be insufficient in particular areas of interest. This is a huge advantage because very specific terminology is utilized in different research areas. For example, the terminology used for objects in Greece is different from the terminology used for North American Plains objects.

However, the ability to add new terms can also be dangerous if it is not controlled to some extent, which it is not in Re:discovery. If anybody can add terms, then the continuity benefit is rendered moot. Users can simply write in the term they would like to use instead of the preset terms.

PastPerfect for Museum Software 4

The final software under consideration in this thesis is the PastPerfect Museum Software 4 (www.museumsoftware.com). Founded in 1996, PastPerfect is a relational database intended to be used by a multitude of museum collections ranging from art to natural history to archives. Archaeological collections are explicitly included among the museum collection types meant to use PastPerfect.

One of the most significant advantages of PastPerfect is that it offers the ability to have multiple collection object types in the same database. Each collection record offers five data entry forms. The user can select an Archaeology, Art, Geology, History, and Natural History. In fact the user can enter data in all five forms for a single record. All five options remain available regardless of use in the record.

PastPerfect offers the option of attaching video and audio clips to a record in addition to other types of files that many other software systems allow to be attached such as a word document, excel file, PDF, JPEG, etc. This allows a better record to be kept of a collection object or site than a few static image files. However, this feature is not free. The ability to attach images costs the user an additional $296-$370 on top of the cost of the basic PastPerfect package (PastPerfect Software Inc. 2009)

It uses a version of Chenhall's Revised Nomenclature for Museum Cataloging for continuity in object description. Chenhall's nomenclature is a widely used nomenclature for many art museums. It has been expanded for description of anthropological and archaeological objects as well. It has the advantage of being a commonly recognized terminological source. It also allows the user to search the Getty Art and Architecture Thesaurus for more terminology and object description.

If the appropriate term is not available, it can be added. While this has the potential for great advantage, it also has potential for abuse as discussed in the section on Re:discovery. It does not allow for categories to be edited. If a term is added to the object description list, the term is not categorized, nor can it be, essentially making the addition useless. In other words, you can add a term for pottery but you can't establish that it is a pottery term or associated with statues not vessels.

Summary of Observations

The database systems currently being used for archaeological collections demonstrated useful characteristics and some areas that can be improved. The first characteristic worthy of inclusion is most of the systems are customizable, allowing the user to adapt the database to his or her needs. The second characteristic worthy of inclusion is allowing users to publish their database information on the Internet, so the rest of the archaeological community (as well as others) could view their data. The third characteristic for inclusion is allowing multimedia attachments. The fourth characteristic is allowing a data dictionary to be incorporated into the database.

However, the databases also showed characteristics to be wary of in designing a collections DBMS. The first characteristic is not being user friendly several of the systems were not easy to set up. Second is being so inclusive that it loses effectiveness. Third is not controlling vocabulary. Some additions may need to be made to vocabulary, but if anyone can add or remove vocabulary entries, it loses all effectiveness. And the final and perhaps most significant characteristic is a lack of interest in recording context. Almost none of the systems evaluated seemed concerned with context or if it was included, it does not allow the specificity needed for context.

Conclusions

Based on the way in which archaeologists are trained to gather information, the way they practice information gathering, the way in which curators, archaeologists, and museum employees use collections management software, and the archeological collections DBMS's currently on the market, a rudimentary understanding of the requirements of an archaeological collections database can be established. The procedures and concerns demonstrated by archaeologists in the field and the texts describing proper archaeological procedures demonstrate that detail in description is extremely important. Everything the archaeologist can describe has to be included because once concluded, the excavation can never be returned to for double checking, except through the notes and records created during excavation. Context is extremely important for archaeology, as Adair (2007) put it, "It is the entire point of an excavation". Archaeologists attempt to document context in as many ways as possible; i.e. writing, pictures, drawings.

The ways in which objects are treated once they are added to a museum collection demonstrate the need to track changes to a record, reflecting changes for or to the object, including changes in location, as well as changes in the understanding of an object, such as the object's original function. Both control and flexibility should be present in object description terminology in order to maintain continuity but still allow for regional differences. Sub numbers, sub-object or sub-accession numbers may be used in order to signal associations between the objects.

The archaeological collections DBMS's currently available demonstrate some requirements are lacking. Despite the concern for context voiced by many archaeologists, it sorely lacks emphasis in many of the collections DBMS's currently available. Most systems currently available offer some form of terminology control or flexibility, but few offer both.

The newfound understanding of the requirements for an archaeological database must be taken into consideration in adapting Specify for archaeological collections. Some of the requirements and concerns voiced by archaeologists will be the same as the concerns of biological collections, but many will be different, requiring changes to the Specify data model. The next step in extending Specify for archaeological collections is to understand how archaeological collections requirements are similar to, and different from, those of biological collections.

Chapter 4 – Comparison of Biological and Archaeological Collections

Now that a basic research information processing requirements for an archaeological collections database have been reviewed, the next step for adapting Specify for archaeological collections is to analyze the significant similarities and differences between those and biological collections. Specify cannot be adapted for archaeological collections without first understanding which elements of the biological collections database management system (DBMS) should remain in Specify and which elements must change in order to meet the requirements of archaeological data concepts and management.

Both biological and archaeological collections management practices have many common requirements for an effective database, that is: 1) to enter, search, and share data 2) to reduce redundancy, and 3) to record administrative and curatorial information. These requirements result in similar needs for comparable relational data tables. However, five critical differences exist between the two collection types. Archaeological collections differ in the: 1) possibility of repatriation of objects, 2) classification methods, 3) object nomenclature, 4) organizational methods, and 5) specificity required for locality description. In a relational database context, those differences will require additional data tables and relationships in a database system.

Similarities

Software designers would have the same basic goal for archaeological and biological database management: to facilitate research and document the administrative and legal aspects of the collection objects. The needs for facilitating research include entering and quickly retrieving data. The needs for documenting the administrative and legal aspects of the collection object include documenting who was involved with the object in each stage of its history, including how it was acquired, legal ownership, the object's location, and data security.

Facilitating Research

The challenge is to provide sufficient data to anticipate and accommodate the needs of both researchers and collections managers. It is impractical to include all available data, since this will often include notes, publications, and other extensive documentation. More will be said in the Differences section of this chapter about what type of data requirements are specific to the collections, but facilitating data entry and retrieval requirement will be discussed briefly below.

The data entry stage is potentially problematic for several reasons, including time and effort required and the potential for mistakes. For these reasons, "The data entry stage is where the majority of projects get stuck" (Sarasan 1981, 47). Detailed information may be required for each collection object. Some of the information may be repeated from object to object or in different tables for the same object. A perfect example of this is a datum for agent (person).

A single agent may be associated with a number of different tables in the database. Take the example of a survey project conducted and documented by a single individual (the agent). An archaeologist excavated the objects, documented the relevant information for them, photographed them, and measured them. Relevant information concerning this agent would have to be entered for each role the archaeologist played as excavator, recorder, identifier, photographer, and measurer.

If the database manager records the archaeologist's name, position, institution, address, and phone number, then five data points need to be entered each time the agent information is added. In this example, the archaeologist played five roles related to the object (excavator, recorder, identifier, photographer, and measurer), yielding 25 data points that need to be entered for just one object. If this is multiplied by 120 objects, this yields 3,000 data points for this basic information. Repeatedly entering this data many times would be an incredible waste of time. It also vastly increases the potential for mistakes, such as misidentifying the agent and in turn not being able to retrieve associated data because the incorrect name was entered.

Searches run on data must be as reliable and comprehensive as possible. If data are entered erroneously, it may be difficult to retrieve. If the data are entered correctly for some but not all data, searches will be incomplete. Ideally, a search should be quickly made and readily comprehended. If one must read through unnecessary verbiage to extract data, valuable time is lost.

Further, data may have to be reentered for a specific analysis. For example, if a researcher wants to find a trend in the rim width of a style of pottery from a certain culture at a particular point in time, she needs to be able to extract the rim widths of all of the pottery fitting the research parameters. If heterogeneous data with mixed semantics is entered in a field, for example, “depth-19cm, rim diameter-27cm, base diameter-12cm, rim width-3cm”, the researcher will have to look at each record and record it in another application, such as Microsoft Excel, in order to use the data. However, if the data are entered in a field by itself, a case of atomization, it can be more easily searched for and used for analysis. It is impossible to anticipate the kinds of data that will be required by every researcher. For this reason, decisions must be made about what data will be entered while leaving open the possibility for the acquisition and recording of additional data.

Administrative and Legal Documentation

Collections management databases serve much the same functions regardless of the collection types for which they are used. Good management requires strict record keeping regarding ownership of the objects in the collection. When museums seek American Association of Museums (AAM) accreditation, they must "legally, ethically, and effectively manage, document, care for, and use the collections" (AAM 2004, 2) and "comply with local, state, and federal laws, codes, and regulations applicable to its facilities, operations, and administration" (AAM 2007). In order to comply with these requirements, administrators must keep or acquire detailed documentation of all aspects of a collection object's life history, from the time it was collected to when it arrived in the collection.

Officials need to know who was involved with the object at all stages, treatments and procedures done to the object, where an object is located at all times, and legal aspects of acquisition and retention. They also need to know who to ask if more information is needed about this. Ideally, a record should provide up-to-date information about who to contact for this. Documentation of agents involved with collection objects also helps with accountability if something is amiss with the object or—in a worst case scenario—if it disappears.

In order to demonstrate effective management and care of a collection, museum administrators must track treatment of an object from the time it entered the collection and also whether special treatment has been required to preserve it. Museum administrators must be able to ensure that proper procedures were followed and to make sure repairs are not confused for original aspects. A significant aspect of documenting the treatment of an object is detailing the condition of the object before and after treatments and/or moves. This documents not only the original condition of the object, but helps one to determine if special treatment is required. Part of the responsibility of being the steward of a collection is being able to locate an object in it quickly (Liu 2008).

Institutions such as the Field Museum use several different means of storing items. Knowing the type of container an object is in and its dimensions can speed up the process of locating it. For example, if it is known that the object in question is in a blue, plastic container with specific dimensions, the person looking for it can eliminate all containers that do not meet those specifications. Search time can decrease dramatically with that simple information. Objects occasionally have a tendency to be moved by one employee unbeknownst to others. Not knowing where to find an object will lead to at best, time lost and at worst, distress over possibly having lost an object (or contributing to its loss or damage).

Use Case Scenario 1 – Location History

National Air and Space Museum, Garber Complex

The National Air and Space Museum storage facility was under the process of assessing and rehousing part of their flight history collection. During this process, they moved the collection objects from their original location to a work table. They took photographs of the objects, assessed their condition, and repackaged them in less-destructive ways. The objects may have been on the work table for anywhere from a few minutes to several days. They subsequently returned the item to its original location, moved it to a treatment area (if required), or moved it to a new, permanent location. In this process, each move of the artifact was recorded in their collection management database. The moves were recorded as soon as the object arrived at its location.

Keeping track of object's temporary location became even more significant when the employees discovered an infestation in an object. Employees assessing WWI officer jackets discovered a carpet beetle under a collar. As a result, the infested jacket and those that had been stored with the infested jacket were placed in polyurethane bags and put into a deep freezer for three days. They were then moved to a Bally box (a heavy, climate controlled container) to warm up for three days. They were then returned to the deep freezer for another three days before being returned to their permanent location. During each phase of the treatment, the temporary location of the object was recorded, enabling the employees to not only keep track of where the object was at any given point but also how long the objects had been kept there. (Miller 2008)

Museum administrators may also need to keep track of where an object had been stored in the past. For example, in Use Case Scenario 1, museum workers had to trace where pestinfested jackets had been stored in order to identify objects that may be at risk and to trace the origin of the infestation.

Being able to locate an object is particularly important for demonstrating ownership. This became readily apparent in the case of the Cosmosphere in Kansas as is demonstrated in Use Case Scenario 2.

Use Case Scenario 2 – Legal Ownership

Kansas Cosmosphere and Space Center, Hutchinson, Kansas

In 2003, the Kansas Cosmosphere and Space Center reported a number of artifacts missing from its collection. The absence was discovered during an inventory of its collection in which it was discovered that there were a number of "irregularities" (Kansas Cosmosphere and Space Center 2005). Upon investigation, the U.S. Justice Department pursued charges of wire fraud, mail fraud, theft, and transportation of stolen property against Max Ary the former director of the Cosmosphere (Pearlman 2005).

In his defense, Mr. Ary claimed that the objects that were sold were objects that he had acquired on his own behalf, some of which were items collected from the trash when the Cosmosphere had thrown them out. He therefore claimed he had full right to sell the objects since they belonged to him, not the museum. He admitted that a few objects that did not belong to him may have been sold accidentally along with the objects that did belong to him (Green 2005). Complicating the prosecution was the fact that ownership documentation of many of the objects was not completed or non-existent. Further, deaccession documentation was not consistently maintained by the museum, so Ary could not prove items he claimed had been "thrown out" by the museum had actually been thrown out rather than stolen (Green 2005).

In late 2005, a jury found Ary guilty believing he knowingly, intentionally sold objects that did not belong to him.

Finally, in order to prove ownership, museum administrators must document transactions involving an object. Ideally, the documentation demonstrates that every change of hands was legal, culminating in the museum's own acquisition of the object. Tables devoted to documenting and tracking administrative and legal issues will therefore be similar or identical between any collection databases. As the Cosmosphere incident demonstrates, thorough documentation of ownership overseen by multiple people could protect objects.

Differences

Despite the similarities between biological and archaeological collections, the significant ways in which they differ affect documentation requirements for each. There are many obvious differences between the two collection types but not all of the differences result in differing database requirements. Only the differences that affect database requirements are relevant to this thesis. Two of the most significant differences are discussed below: 1) how the collections are perceived and 2) the role of locality. The approaches to them are demonstrated in three ways: 1) classification method, 2) nomenclature, and 3) organization method. The role of locality is demonstrated through the specificity required of locality information.

Perception of the Collection

The differences between biological and archaeological collections include the ways in which people in the field perceive them. These directly affect the ways the collections are classified, the nomenclatures used, and the ways they are organized.

Method of Classification

Classification of objects in a collection makes it more accessible and more valuable for research. Classification attempts to group objects together in recognizable, functional manners, making their understanding and use easier (Adams and Adams 1991). Classification affects and is affected by nomenclature and organization. While biological collections benefit from a common classification system, one does not exist for archaeological collections due to a lack of consensus.

Biological specimens are named based on a formal nomenclatural protocol system derived from the time and work of Carl Linnaeus in the 1700s. Biological classifications are imperfect, but they have as their goal a taxonomy that is generally and widely accepted, derived from sound theoretical and historical principles based on patterns of decent and relatedness.

Archaeology lacks a comparable legacy and protocols for the subject classification of objects, although many have been attempted. There are only "hodge-podge approaches about classifications" (Odess 2007). These have produced a multitude of different classification methods, due in large part to the quirks of historical processes and lack of consensus regarding criteria due to variations in research objectives.

Use Case Scenario 3 – Method of Classification

Archaeology Lab, Kansas Historical Society

The Archaeology Lab at the Kansas Historical Society (KHS) has compiled a complex, detailed taxonomic tree for the artifacts in their collection, created in large part due to the need for accuracy and standardization in their records. Much of the data entry is performed by volunteers and interns, not all of whom have a great deal of experience with artifacts. Even when they have experience, the terminologies that they have used before may not be the same used at the KHS.

Artifact classification is based almost entirely on function. The data dictionary used is first separated into general areas of use. For example, there are categories for AMMUNITION, AUTOMOTIVE, BY PRODUCTS, CLOTHING, COMMUNICATION, CONSTRUCTION, etc. The classification is then broken into function. For example the general area FURNITURE AND APPLIANCES, is broken into FURNITURE, MAJOR APPLIANCES, MINOR APPLIANCES, CLEANING APPLIANCES, FURNISHINGS, etc. This is followed by type. For example, FURNITURE is broken into CHAIR, SOFA, TABLE, DESK, etc. After type is subtype, i.e. for chair; folding, rocking, straight back, etc. Then the portion of the subtype, such as complete, leg, seat, etc is recorded. All of which is intended to allow the records to be as precise and accurate as possible

Unlike biological classification, archaeological classification and typology is almost entirely time-, space-, culture-, material-, style- and archaeologist-dependent. This has resulted in numerous classificatory systems, many of which are institution specific. For example, the archaeological division of the Sam Noble Oklahoma Museum of Natural History at the University of Oklahoma and the Kansas Historical Society use different taxonomic trees that they themselves created (Garst 2007, Leith 2008).

The institution-specific classification phenomenon is due in large part to the belief that:

> … typological concepts have no fixed or inherent meaning apart from their use, which varies from typology to typology and from person to person. It is therefore impossible to discuss types and typologies except in subjective terms. We cannot speak of *the* concepts; we can only speak of our concepts (Adams and Adams 1991).

In other words, the research questions and paradigms of the archaeologists have a great impact on the classification used. Since those questions and paradigms are often not the same, and certainly not universal, attempts at a universal classification have been severely complicated. However, the task has not been abandoned.

Furthermore, the type of material being classified impacts the method used, thus the approach to classification may differ from object to object. For example, the classification of pottery in the American Northwest has given rise to a number of complex classification methods (Adams and Adams 1991). Pottery analysis in the American Southwest creates types with different methods (Adams and Adams 1991). In another example, Nubian Pottery and Egyptian wares are classified completely differently though they are found in the same sites (Adams 1986). Other methods, such as Robert Chenhall's system based the proposition that "every man-made object was originally created to fulfill some function or purpose." (Chenhall 1978) or the Getty Museum's Art and Architecture Thesaurus, have produced more widely used classifications. Though widely used, these fall short of the universality of the Linnaean taxonomic system in biological classification.

The Kansas State Historical Society, for example, uses classification based on functionality. Functions of objects in their historical collection can be known from oral history, written accounts, pictures, and current use. However, descriptions are not available for many objects. In the cases of objects from cultures that did not have writing, existed millennia ago, and left no indication of how objects were used—and for which archaeologists have not been able to determine function--this information may not be available. In some cases, an object's function may be the object of speculation, subject to change in light of new data and theories. In the case of objects with multiple functions, some may be given priority over others and not all may be used in a given classification.

Form is also used in classification. It is considered by some to be a more reliable category than function because it is the only aspect of the artifact that can be known by archaeologists (Adams and Adams 1991). However, identification of form becomes much more difficult in the case of pottery shards of a vessel for which the complete form is unknown.

Classification by either function or form is highly dependent on nomenclature. If a classification system is to be used by multiple institutions, the nomenclature must be agreed upon and shared.

Nomenclature

Nomenclature describes and defines objects while classification systems organize the named objects into recognizable patterns (Rickard 1944). Creating a common nomenclature is an essential part of creating a classification system. Chenhall, for example, is primarily concerned with creating a common nomenclature with classification as a secondary concern. In order to classify an item, the archaeologist must know what to call it and a common nomenclature is essential if there is to be a common classification system.

Similar to the classification issue, while biological collections have benefitted for centuries from a disciplinary consensus regarding the goal of building a common nomenclature, this has not occurred for archaeological collections. This is due neither to a lack of acknowledging its importance nor a lack of trying. However, attempts have not been successful.

The United Kingdom chapter of the Computer Applications and Quantitative Methods in Archaeology (CAA), an international association for both archaeologists and computer scientists, stated that a common nomenclature is necessary for sharing information within an institution it is essential to functionality and data cannot be shared intra-institutionally without it (Gardin 2002). Currently, several efforts funded by the National Science Foundation are being made for standardization of nomenclature in the form of an Encyclopedia of Archaeology. One example is a project involving Dr. Mary Adair at the University of Kansas, whose project is directed specifically towards the standardization of archaeological and anthropological nomenclature for database use and sharing (Adair 2007).

Some institutions, such as the Getty Museum has created an Art and Architecture Thesaurus (AAT) that proposes a nomenclature and hierarchy for man-made objects, including archaeological objects. For example, AAT defines the term "amphorae" as "Ancient Greek and Roman storage vessels of many variations usually having a large oval body with a narrow neck and two or more handles extending from the mouth or neck to the shoulders on the body" (Getty Art and Architecture Thesaurus 2004). It further offers a hierarchy for the term within OBJECT FACET> FURNISHINGS AND EQUIPMENT> CONTAINERS> CONTAINERS> STORAGE VESSELS>

AMPHORAE. While this work is widely used and referenced, it still falls far short of the thoroughness and universality of Linnaean taxonomy.

Method of Organization

Differences in classification are reflected in how a collection is organized and in turn are reflected in database organization. Biological collections are separated into their taxonomic units while archaeological collections are usually conceived of in terms of geography, objects, and sites rather than types. Biological departments are each distinct, homogenous units often, though not exclusively, corresponding to a single category of biological organism (Bentley 2008).

In museums with biological collections, each different collection type often has its own database (Garneau 2008). For example, in the Museum of Natural History at the University of Kansas, there are separate divisions and databases for herpetology, mammology, ornithology, botany, etc. By contrast, an entire archaeological collection, such as those of the Field Museum in Chicago, the Kansas Historical Society, and the Yale Peabody Museum, may be cataloged in a single database rather than separating objects into divisions as in biological collections.

Even when objects are separated into different collections or divisions, many different types of objects may be included together. For biological collections, each database has its own distinct set of data requirements for the discipline documented in it. While many of the data requirements may be the same, requirements for object attributes and collecting attributes may differ. For example, the attribute data requirements for a mammal collection (see Figure 1) may be different from ichthyology (see Figure 2).

Figure 1 Screen Shot of the Collection Object Attributes table of the Mammal collection at the Biodiversity Research Institute at the University of Kansas (using Specify).

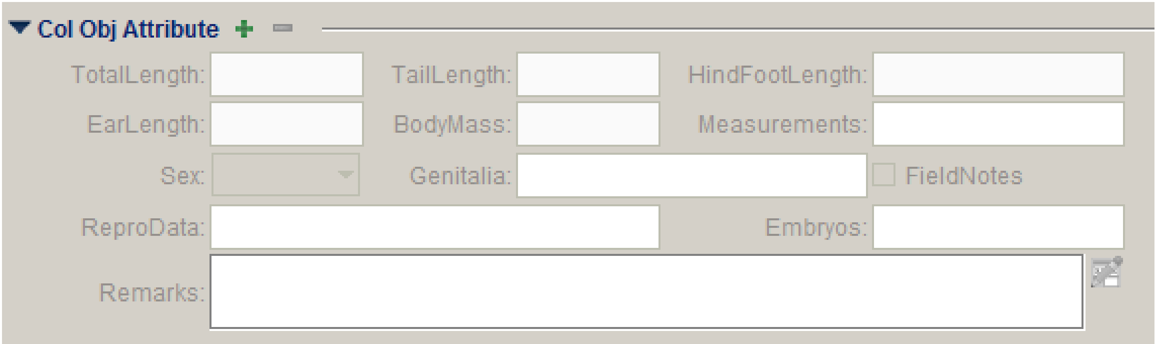
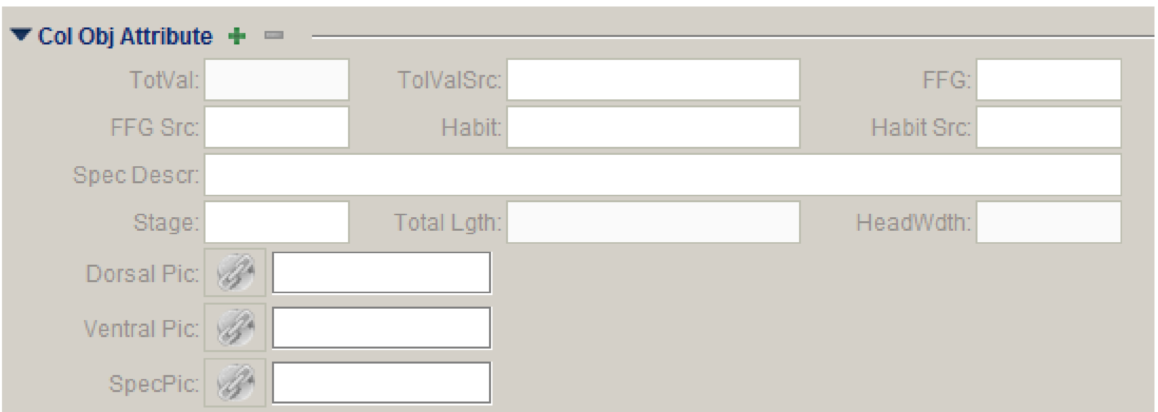


Figure 2 Screen Shot of the Collection Object Attributes table of the Aquatic Insect collection at South Dakota State University (also using Specify).



Archaeological collections are typically not divided into separate distinct units corresponding to object types (i.e. ceramics, lithics, faunal remains, etc.) However, a biological collection will often be comprised entirely of one class of organism with the same data requirements (age, sex, beak length, wing span, body length, etc.); an archaeological collection will include many different types of objects that have significantly different data requirements.

Archaeological collections usually prioritize geography. A large part of the significance of an object derives from where it was initially collected. Thus, a museum's archaeological collection (such as those of the Field Museum, Yale Peabody Museum, the Kansas Historical Society, and the University of Kansas Archaeological Research Center) is often divided based on locations such as continent, region, and site with groupings of different object categories.

Archaeological collections, though generally kept within one department sharing the same database, they are often physically separated according to geography (when storage space allows). For example, the Kansas Historical Society separates objects according to site. Within the site, the objects may be organized or reorganized according to the needs of the researcher using the collection (Garst 2007). Thus many museums, such as the Field Museum and the Smithsonian, have divisions by regional areas, such as Americas/New World, Europe, Asia, Oceania, Africa. The collections may then be further divided depending on the institution. The Field Museum divides its collections based on CONTINENT then COUNTRY then CULTURE. They also have a separate division for human remains. Human Remains work in association with all other divisions (Tilson 2008).

Use Case Scenario 4 – Method of Organization

Field Museum

The Field Museum separates its collection of over half a million artifacts into sub-collections that are based almost entirely on geography. In fact, the collections are organized into geographic area before they are divided into disciplinary area. The collection is not divided into "archaeological" or "ethnographic" until after it has been divided into geographic region.

The sub-collections are Africa, Asia, Europe, Human Remains, Mesoamerica and South America, North America, Pacific, but also include Malvina Hoffman Collection, and Exhibitions. If the sub-collection is large enough, it is then divided further into parts of either a more specific geographic region, discipline, or both. For example, Mesoamerica and South America are divided by Mexican Ethnography, Mexican Archaeology and Maya, Lower Central America/South America (CAM/SAM) archaeology, Lower CAM/SAM ethnography, etc. The collections are then divided by Curator (who may be responsible for multiple parts in multiple sub-collections). For example, one curator is responsible for the categories of North American Ethnography, North American Archaeology, and South American Colombian Gold (Tilson 2008).

An exception to this classification method is donor or teaching collections. Donor collections are a set of objects donated by a single donor. Teaching collections are a collection of objects used by the museum to show to museum patrons or institutional students. Organizations such as the University of Kansas, the Field Museum, and the Yale Peabody Museum keep their archaeological collections together, regardless of the collection's contents or the geographic origin of the objects. These collections offer the same database problem as the rest of the archaeological collection; they frequently contain a mix of object types.

Unlike biological collection databases, archaeological collection databases contain all of the collection object information in the same database. Because of this, an archaeological collection database design must allow for different distinct sets of data in the same database. The database may need to account for ceramic statues and ceramic vessels, metal pins and metal plates, stone projectile points and stone stelae.

Role of Locality

The second and possibly the most significant difference between biological collections and archaeological collections is the importance of the specific location—a much smaller unit than general geographic area--where an object was collected. Database requirements differ greatly from those for biological collections due to the specificity and range of data required.

Importance of Locality

For biological collections, the specific locality where a specimen was collected is useful but not necessary for much of species analyses. For archaeology, locality, which is often essential for a successful analysis, is of the utmost importance. Without locality, an object's analytical value is limited.

While biological collections databases can record locality, it tends not to factor heavily in analysis of the material. A clear exception to this generalization is in biodiversity studies, which combines focuses on environment and organisms in order to understand the relationship between the two. However, even in this branch of biology, the concern with locality is not as specific as in archaeology (Morris 2005). That is, the precise location at which a specimen was collected is usually not required for analysis.

The greater concern for biological collections objects analysis is the habitat in which the collection object was obtained (Bentley 2008). Location is important for logistical reasons, such as knowing where to go to collect additional specimens. Biological collections such as those at the University of Kansas are concerned with the natural features of the area in which the specimen was collected. They need to know roughly how far from a river a specimen was found, or in the case of ichthyology, how far from shore, the depth, the water type, etc.

Archaeology, on the other hand, is locality-dependent. From the general (country and site) down to the very specific (layer, feature, provenience), locality is vital. Ideally, the provenience of each object is carefully recorded (Banning 2000). Preferably, provenience includes not just the country, state or province, and site designation, but also the precise spatial location (latitude and longitude or UTM coordinates) and the location of the object relative to a site-specific datum and the depth at which an object was excavated. As described in the third chapter, archaeologists attempt to document information about the precise provenience and matrix in which an object was discovered in as much detail as possible. For example, most archaeologists do not just document measurements with written numbers but through drawings, photographs, and verbal descriptions.

Locality concerns deal not just with where an object was found but what was found in association with it. In archaeology, context and associations are of paramount importance. The specific, detailed spatial relationships of objects to one another and to their contexts are often critical for analysis. Often, these relationships can only be preserved in documentation. A carefully made set of illustrations, such as level plans, profiles, drawings depicting its provenience is often used more to depict the context of the object than the appearance of the object itself.

An archaeologist ideally needs to know what objects were found next to each other, which were found in the same context, which were found in the same assemblage (site or pit depending on how the excavators decide to divide the objects), which were found in the same layer, etc. The exact nature of the association of the objects is significant as well (to the east, one inside another, one above another, etc.). The placement of one object in relation to another, its distance from it, etc. is valuable information. If this is not recorded, much of the most valuable data about an object are lost.

Biologists may be interested in the physical environment of a collection object, but most current studies are not interested in the spatial relationships between one fish and another and do not record it (Bentley 2008). If species interaction is recorded, the information tends to be general i.e. "fish one" was found attached to "fish two" or in the vicinity of "fish three." The exact association of the fish, i.e. "fish one" was captured 3 cm to the north of "fish two," is not usually a part of the methodology. Nor are the methods of collection, such as netting, conducive to knowing how the animals were placed before the collection.

In archaeology, the feature with which an object is associated is significant as well. As described in Chapter 3, in archaeology, a feature are evidence of human actions that cannot be moved (and therefore cannot be collected). A feature provides another level for describing locality and context. It presents special challenges for a database system since a collected object can be documented with reference to the features in which or near which it was found. The requirements and challenges for features will be discussed more in depth in the next chapter.

Stratigraphy, the observable layers caused by the accumulation of natural or cultural deposits over time (Renfrew & Bahn 2004), is one aspect of locality in archaeology. The stratigraphy used by biological collections and archaeological collections have some similarities but have significant differences as well. Paleontological collections are the only biological collections for which stratigraphy is recorded. They use two primary types of stratigraphy, both of which offer different information. The first type is lithostratigraphy, which is concerned with the type of rock in which an object was collected (Farrell 2011), which gives insight into the environment of the area at the time of deposition. The second type of stratigraphy is chronostratigraphy, which records the time the matrix in which an object is found was deposited (Farrell 2011). Both are treated as hierarchies.

For archaeology, stratigraphy can indicate a relative date for objects. It also gives a contextual reference. Stratigraphy may vary from site to site but also within a site. If there are multiple excavation pits in a single site, each pit may observe and record a different stratigraphy. Archaeological stratigraphy borrows the geologic principles for stratigraphy, but adapts them for archaeological purposes. For example, stratigraphy is not considered a hierarchy in archaeology. Harris points out two reasons why archaeology cannot directly copy the geologic principles 1) archaeological stratigraphy is almost always the result of human interaction and not directly subject to geologic principles 2) archaeological objects are created, preserved, or destroyed by human action and therefore not subject to natural rules (Harris 1989).

Harris proposed four laws for archaeological stratigraphy, adapting the three geologic principles and adding a fourth. The laws he proposed are: 1) the Law of Superposition, 2) the

Law of Original Horizontality, 3) the Law of Original Continuity, and 4) the Law of Stratigraphical Succession. Both archaeology and paleontology follow three geologic principles: the Law of Superposition, the Law of Original Horizontality, and the Law of Original Continuity (Harris 1989).

The Law of Superposition states that lower layers are older than higher layers. The Law of Original Horizontality states that the strata were originally deposited horizontally. If the strata are no longer horizontal, the change in position occurred after deposition. The Law of Original Continuity states that strata are formed in one piece and the edges should feather out. If the strata is broken up or has sharp edges, it occurred after deposition (Harris 1989). Harris added the proviso "as originally created" lower levels are older than higher levels, acknowledging that humans may have broken through lower layers resulting in younger features in older layers (Harris 1989, 30).

For the archaeological Law of Original Horizontality, Harris (1989) modifies the law to say strata will tend to be horizontal. If the stratum is not horizontal, it was either deposited that way either as a result of human action or because of the contours in the strata below it (Harris 1989). For the archaeological Law of Original Continuity, Harris modifies the law to say strata will be bound by the features of the landscape or will feather out at the edges. If a stratum has a sharp edge that is not the result of an environmental boundary, it occurred after deposition (Harris 1989). The final law is the Law of Stratigraphical Succession. This law states that a stratigraphic unit takes its place from its position between the units above it and below it and with which it has physical contact (Harris 1989).

Stratigraphy is vital for establishing association and chronological relationships of objects from a site. In the early years of archaeology, stratigraphy was the only means of establishing change over time at a site. Although today there are other means of determining object dates (radiocarbon dating, etc.), stratigraphy remains one of the most significant means of documenting the chronological association of objects.

Summary of Differences

Archaeological and biological collections differ in several significant ways, resulting in different database requirements. Two of the most significant ways in which the two collection types differ is the way in which the collections are perceived and the role of locality. Perception of the collections is broken into three parts: how the collection is classified, the use of common nomenclature, and how the collection is organized. While biological collections are more universally standardized in their methods, the archaeological collections are more segmented.

Possibly the greatest difference between the two collection types is in the way Locality is treated. Archaeologists require the geographic location in which an object is discovered to be described in detail, including context, and may require tight security on the details due to the threat of looting. Biological collections, on the other hand, collect locality data but not to the extent and amount of detail required of archaeological collections. An archaeological database must allow the archaeologist to record not only the site at which an object was discovered and the location information for the site, but also the exact location at which an object was excavated and the spatial relationships it held with other objects and, if images are available. It must also allow the database user to view level plans, profiles, drawings, and photographs documenting these relationships (Lock 2003).

Conclusion

A basic understanding of what needs to be done to adapt Specify for archaeological collections emerges from an analysis of the similarities and differences between archaeological and biological collections. Because Specify programmers would have similar goals for both collection types, *i.e.* facilitating research and documenting the administrative and legal aspects of the collection objects, some of the tables in the database model will remain unchanged.

However, several significant differences exist between the two database types that require changes to the underlying model information. Two of the most significant differences, perception of collection and role of locality, are essential to the make up of those database types. The differences in the areas of method of organization, nomenclature, method of classification, and the importance of locality, result in different database requirements for the archaeological database. Each of these differences will have a significant effect on the database model.

Now that a basic understanding of the similarities and differences between the database requirements for archaeological and biological collections is understood, we can begin to adapt Specify's data model for archaeological collections. The details of how this can be accomplished will be discussed in detail in the following chapter.

Chapter 5 – A Model for Extending Specify for Use with Archaeological Collections

Now that the similarities and differences between archaeological and a biological collections have been reviewed, the next step for adapting Specify for archaeological collections is to determine which tables in the data model should stay untouched, which tables should stay but must be modified, and which tables should be added in order to accommodate similarities and differences between the collections types as discussed in the precious chapter.

As noted previously, biological and archaeological collections have many of the same basic requirements with regard to creating effective databases that facilitate research and document administrative and legal information. These similar requirements result in the need for similar data tables in a database system. Most of the data tables that already exist in Specify can therefore be carried over into a database for archaeological collections. However, as noted, there are also critical differences between the two collection types resulting in the need for new data tables to be added to the database in order to adequately address the archaeological data concepts. Archaeological collections need to be able to document a variety of collection object types with varying sets of attributes for each. They also require more detail and description pertaining to the excavation locality of a given collection object, including highly specific information. To meet this need, tables were added to the relational database proposed here to document a collection object's provenience, associated feature, and context.

Recycled Approaches and Tables

The similarities between biological and archaeological collections are in the areas of: 1) facilitating data entry, 2) data retrieval, and 3) documenting administrative and legal aspects of a collection object; including documenting agents, past owners, acquisition, and object location. These entail the use of similar database tables and fields as will be discussed below. Some tables and approaches that already exist in Specify can be recycled for an archaeological database, with little or no modification.

Facilitating Research

Facilitating research is a prime goal of most collections databases. Two aspects of research facilitation will be discussed below: data entry and data retrieval. These aspects are accommodated through the use of a relational data model with atomization. Both already exist in the Specify model and can be beneficial to carry forward for archaeological collections.

The use of a relational data model resolves much of the concern for data entry. In a relational model, tables are in some way related to other tables in the model; none are isolated. A datum can be accessed by tables other than the one in which it is stored. Data only has to be entered in one table to be used by another that may need the data. It offers a single point of entry for any point of data.

To demonstrate this more concretely, we can look at the example from the previous chapter: entering agent-generated data for a survey project. If a single person were responsible for multiple aspects of an object's collection and documentation, entering the relevant data for that person could require 3,000 points of entry. However, if the database is relational, 3,000 entries quickly drop down to five.

Instead of entering an agent's information such as an archaeologist's into 1) the PROJECT[2] table, 2) the PHOTO table, and 3) twice in the COLLECTION OBJECT table, her information is entered only once in the AGENT table, producing a single AGENT record. The AGENT record contains all five of the data points. A link to that AGENT record can be put into the PROJECT,

[2] Small caps were used to identify concepts that refer to concepts within the Specify data model.

PHOTO, and COLLECTION OBJECT tables so the agent information can still be available on each of the tables, without reentering it.

Single-point entry addresses the problem of having records with mismatched information for the same agent. When there are multiple records for the same agent, there is not only the chance that the data can be misentered at some point, that the records may become incongruent. As was demonstrated in a previous chapter, there could easily be 3,000 or more records that would have to be corrected. With single-point entry, only one record would have to be corrected. However, because there is only one record, the need to accurately enter data increases. If the data are entered incorrectly, it will be incorrect for all linked records.

Data Retrieval

The issue of quick and easy data retrieval can be resolved through *atomization*, entering each piece of data into a separate field (Morris 2005). Data entered in separate fields can be more easily searched for and used in analysis. For example, a researcher may have to find ceramic vessels from burial contexts manufactured using coiling with grey clay from the Valdivia culture (3,000 - 2,500 BC). Without atomization, the researcher may have to search through potentially hundreds of thousands of individual objects, sorting out the ones that are ceramic, then sorting out the ones that are vessels, then sorting out the ones made of coil wrap technology, etc. to find the five objects that fit the multiple criteria. However, if atomization is used, a query can use specific attributes to quickly narrow the search.

Atomization requires tradeoffs in practicality. In an archaeological collection, the information required for each object is so diverse that individual fields for every possible datum—if these could even be known--can burden data entry and computer memory to the point of severely reducing database functionality. A database must represent a balance between the system's efficiency and having enough fields to reasonably document the collection object information. Because new data may be entered in the future, it must also be flexible enough to accommodate unanticipated needs. It is impractical to assign an individual field to every possible datum because there may be an infinite number of these. Research objectives change from researcher to researcher and over time. Although good guesses can be made based on current and projected use, it is impossible to predict all the data that will interest present and future researchers. A possible solution for this is discussed in the New and Modified Tables section of this chapter.

Administrative and Legal Aspects of Documentation

Biological and archaeological collections are also similar with respect to administrative and legal documentation. Tables that meet these requirements can be arranged into four groups: 1) tables for general documentation, 2) tables for condition documentation, 3) tables to track transactions, and 4) tables for object location (which includes tables for EXHIBIT HISTORY, PROVENANCE, STORAGE CONTAINER, and STORAGE LOCATION). Most of these tables are currently in Specify and no modification is necessary. However, some slight modifications and additions are required for some tables.

General Documentation

General documentation is documentation of general information, including who has been involved with collection objects, their addresses, institution information, and institutional agreements. For general documentation, Specify currently uses tables entitled; ADDRESS, AGENT, INSTITUTION, PERMIT, and REPOSITORY AGREEMENT.   For a more concrete example of this, consider the example of an institution acting as a repository for another institution. The REPOSITORY AGREEMENT table allows the collection manager to document information: when the agreement is made, the agreement number, who authorized the agreement, who received the object, whether the object is currently being held by the institution, and when the object is returned. The AGENT table allows the collection manager to document information about the people involved in the agreement such as their name, title, email address, and position. The ADDRESS table allows the collection manager to document the physical address for the agents involved as well as the institution. These do not have to be amended for archaeological collections.

Condition Documentation

Condition documentation is documentation of the state of an object and changes to it. For condition documentation, Specify currently uses tables such as; APPRAISAL, CONSERVATION DESCRIPTION, CONSERVATION EVENT, and DNA SEQUENCING. For a concrete example of this, consider the example of a metal axe head found at a Pawnee site.

The CONSERVATION DESCRIPTION table allows the collection manager to describe the axe head, the conditions the axe head should be kept in, and the treatments it should receive. The CONSERVATION EVENT table allows the collection manager to document that the axe head was cleaned and received electrolysis to remove rust. The table can record who authorized the treatment, who performed the treatment, when it was performed, and any notes on the treatment process.

Although DNA Sequencing seems like a uniquely biological need, it is also valuable for archaeological collections. As mentioned earlier, archaeological collections contain different object types, including animal, human, and agricultural remains. DNA analysis has become another significant tool for archaeologists (Renfrew and Bahn 2004). DNA sequences allow the archaeologist to describe the organic material with a level of specificity that no other table can accomplish.

The CONSERVATION DESCRIPTION, CONSERVATION EVENT, and DNA SEQUENCING tables allow the user to document any changes made to an archaeological collection object. An archaeologist is ethically bound to document any modifications made to an object in order to prevent others from believing the modern changes are original (Liu 2008). All of these tables are sufficient to satisfy the needs of an archaeological collection and do not need to be amended.

Tracking Transactions

"Transactions" includes any transference of ownership or stewardship of an object. Specify uses tables such as; ACCESSION, BORROW, DEACCESSION, EXCHANGE IN, EXCHANGE OUT, GIFT, LOAN, and SHIPMENT to track transactions. Each of these tables documents the change of ownership either to the institution, as in the case of ACCESSION, BORROW, and EXCHANGE IN, or from the institution as in the case of DEACCESSION, EXCHANGE OUT, GIFT, and LOAN. Although not documenting the change of ownership, the SHIPMENT table is grouped with these tables because it documents the transition of collection objects from one institution to another. These tables are advantageous for an archaeological collection, but two more tables must be included to account for the need of some archaeological collections: 1) objects to be repatriated and 2) the need to document past ownership.

The North American Graves Protection and Repatriation Act (NAGPRA), put into effect in 1990, requires that any institution receiving Federal aid return Native American human remains and sacred, cultural objects to the tribe to which they originally belonged, if it can be identified (Renfrew and Bahn 2004). A unique table was added to the data model to allow for the documentation of archaeological collection objects in accordance with this law. Although repatriation is a type of deaccession, it has requirements separate from that of a DEACCESSION table.

NAGPRA requires documentation of every aspect of the return of the remains and objects. The Repatriation table includes fields to document the federal register notice information, the date the notice was published, which tribe is claiming the remains or objects, when the claim was made, what the resolution to the repatriation effort was, and whether it is complete or not.

Repatriation has a many-to-many relationship with Deaccession, which in turn has a one-to-many relationship to Collection Object. This relationship means that one repatriation act may include multiple collection objects. In one act, remains and multiple objects may be returned. The remains and each of the objects involved may be considered a separate collection object, each with its own catalog number.

This relationship also means that a collection object may have multiple repatriation acts. Repatriation includes incomplete repatriation as well as completed repatriation. By this I mean, repatriation efforts may be started, but may not result in deaccessioning the objects. These efforts still need to be documented, even though they did not result in repatriation. The same object may be the subject of another repatriation effort later. The many-to-many relationship allows both efforts to be documented.

Figure 3 Repatriation Table

Repatriation	
PK	**Repatriation ID**
	Date of Request
	Requesting Tribe
	Requesting Agent
	Investigating Agent
	Resolution
	Complete?
	Timestamp Created
	Timestamp Modified
FK1	Deaccession ID
	List Number

Repatriation also has a one-to-many relationship with Shipment. This relationship means that one repatriation effort can be linked to multiple shipments. This relationship is used because multiple collection objects can be associated with a single repatriation effort. The different collection objects may be shipped in multiple packages in multiple ways. One object may be handed over to a representative while others were shipped. Both of these methods need to be documented.

NAGPRA requires documentation of not just the deaccession itself, but any attempts to discover to whom the remains or objects in question may belong. This includes publishing notice in the federal register as mentioned above and any consultations with people regarding ownership of the remains or objects.

In order to accommodate the documentation of consultations, I added a table called Repatriation Agent. This table has a many-to-one relationship with repatriation. This relationship means that there may be multiple consultations for each repatriation act. This table includes fields for who was involved in the consultation and when the consultation took place.

Figure 4 Repatriation Table

Repatriation Agent	
PK	**Repartiation Agent ID**
	Role
	Remarks
	Contact Date
	Timestamp Created
	Timestamp Modified
	Agent ID
FK1	Repatriation ID

Repatriation documents the transfer of ownership, or attempted transfer of ownership of collection objects that are being deaccessioned from the collection. For objects in the collection, the Provenance, defined as the ownership history of an object, needs to be documented. Provenance is another means of documenting an object's history.

Provenance is intended to demonstrate that an object passed legally from hand to hand until arriving at the institution in which it resides. The Provenance table also allows for the condition of the object to be tracked. The fields Condition Received and Alterations Performed were added to allow documentation of changes in the condition of an object during the time of ownership. A link to Conservator Description is also present for a more detailed documentation of condition if desired.

Figure 5 Provenance Table

Provenance	
PK	**<u>Provenance ID</u>**
	Owner
	Dates Owned
	Aquisition Method
	Condition Received
	Alterations Performed
	Timestamp Created
	Timestamp Modified
FK1	Agent ID
FK2	Conservator Description
FK3	Artifact ID

Provenance has a "many-to-many" relationship with Collection Object. This means multiple objects may have many provenances and a single provenance may involve multiple objects. For example, a set of paired Sun and Moon statues may have been found together but purchased separately by a series of organizations before ultimately being purchased and reunited by a single museum. In this case, there would be multiple provenances for each, one would be the same but the others would be different.

Object Location

Documentation of object location includes information related to where the object is and where it has been. Two tables are used in the archaeological data model to achieve this goal; STORAGE LOCATION and EXHIBIT HISTORY. The STORAGE LOCATION table is currently in the Specify data model and although it is valuable for an archaeological collection, it is insufficient to document an object's location history. The STORAGE LOCATION table was modified by adding fields and its relationship to COLLECTION OBJECT and a new table, EXHIBIT HISTORY, was added in order to adequately document a collection object's history. Although the addition of the STORAGE LOCATION and EXHIBIT HISTORY tables are suggested for archaeological collections, they would be beneficial for biological collections as well.

To document the location of an object, Specify currently uses a table called STORAGE LOCATION which is organized in a storage tree. The storage tree allows the users to document where an object is in an efficient manner, by making storage location into a tree. Making the storage location into a tree allows the user to type in part of a storage location, such as "Shelf B", then choose the correct shelf from the options returned instead of typing in the building name, collection, floor, room, aisle, cabinet, and shelf in order to say where the collection object is stored.

The STORAGE LOCATION table is intended to document where a collection object is stored. Currently, STORAGE LOCATION is a child of the PREPARATION table, meaning that storage location is associated to a collection object through the collection object's preparation (i.e. skeleton, x-ray, ethanol, cleared and stained, etc.). Only one storage location can be assigned to a collection object preparation. Past locations or location changes cannot be tracked. The storage table must allow multiple storage locations to be associated with a collection object, only one of which is current.

For this reason, the relationship between Storage Location and the collection object must be changed for an archaeological collection. Archaeological collections will not require a PREPARATION table as is used in biological collections. Although Storage Location is linked to the COLLECTION OBJECT through the PREPARATION table for biological collections, Storage Location is linked directly to the COLLECTION OBJECT table in the archaeological model.

Instead of being a "many-to-one" relationship between PREPARATION and STORAGE LOCATION, this has been changed to a "many-to-many" relationship between COLLECTION OBJECT and Storage Location. This relationship allows more than one collection object to be associated with a single storage location and allows a collection object to be associated with more than one storage location. This change means a collection object can be associated with a current location, a temporary location, a treatment location, and an exhibit location, if appropriate.

Figure 6 Storage Location Table

Storage Location	
PK	**<u>Location ID</u>**
	Location Name Current? Date Placed Date Moved Text 1 Text 2 Text 3 Text 4 Number 1 Number 2 **Timestamp Created** Timestamp Modified
FK1	Artifact ID

A field called "Is Current" was added to the data model for archaeology in order to document the current storage location. The Is Current field allows the user to modify locations as the object location changes. This ability is particularly useful in situations in which users need to reconstruct where an object had been stored in the past. For example, in Use Case Scenario 1 (in the previous chapter), museum workers had to trace where infested jackets had been stored in the past in order to identify other objects that may be at risk and attempt to trace the origin of the infestation. Because they were able to trace the location history of the objects, they were able to identify at-risk objects and treat them for possible infestation.

I added a table for Storage Container as a more specific way to document in what the object is stored. The fields are intended to inform the user what to look for when they try to locate the item. As discussed in the previous chapter, knowing the type of container an object is in and its dimensions can speed up the process of locating an object. For example, knowing that the object is in a small, blue box with code 14-b within a large, brown box with code SB-10, will allow the user to locate the desired object much quicker than having to search through every container in the area.

Figure 7 Storage Container Table

Storage Container	
PK	**Storage Container ID**
	Container Type
	Description
	Name
	Number
	Bar Code
FK1	Location ID
FK2	Artifact ID
	Barcode ID
	Timestamp Created
	Timestamp Modified

I added three means of labeling the container in the Storage Container table, Code, Barcode, and Electronic code. Different institutions use different means of labeling containers (if they are labeled at all) and may even use several means if they are switching from one tracking method to another, or started switching in the past and never completed the switch. For example, the Field Museum uses several means of labeling their containers. They started with a code on each container and have started moving to barcodes (Tilson 2008). A code may consist of text, numbers, or a combination of the two.

EXHIBIT HISTORY is also valuable for tracking location history. When an object is put on exhibit, its location is often changed, either from storage to an exhibit space or from one museum to another. Tracking exhibit history allows people to track not only where an object is or has been but also the objects to which it has been exposed, fulfilling the need to know an object's location at all times and potential of contamination, as described above and in the previous chapter.

The fields in the EXHIBIT HISTORY table are intended to track the location of an object and the objects to which it was exposed during the exhibit. The fields Exhibition Start Date and Exhibition End Date track the time of the exhibition. These fields along with Location, Problems, and # Pieces, help users to be able to track a collection object's exposure to other objects. The field "Permanent?" is present for those objects that are on permanent exhibit, rather than a temporary exhibit (Miller 2008). The other fields, Exhibition Title, Curator, Description, and Remarks, were added to help the user identify the exhibit and document accountability (Curator).

EXHIBIT HISTORY has a "many-to-many" relationship with COLLECTION OBJECT. This relationship means there may be multiple objects in one exhibit and an object may be in multiple exhibits. In this way, museum officials may be able to track the history of exposure of an object to other objects and narrow the timeframe for the exposure. The field Problems allows officials to document if any issues were identified with the exhibit, for potential contamination and treatment.

EXHIBIT HISTORY has a "many-to-many" relationship with LOAN. This relationship means that there may be multiple objects on one loan (for example if three objects were loaned to another institution for an exhibition) and one object may have been loaned out multiple times. This relationship was put in place in case an object is loaned to another institution for exhibit. If it is, a record still needs to be kept of the location of the object, the duration of time spent at the location, and, if possible, the other objects on loan with it.

Figure 8 Exhibit History Table

Exhibit History	
PK	**<u>Exhibit History ID</u>**
FK1	Artifact ID
	Exhibition Title
	Exhibition Start Date
	Exhibition End Date
	Location
	Curator
	# Pieces
	Permanent?
	Description
	Problems
	Remarks
	Timestamp Created
	Timestamp Modified
FK2	Provenance ID
FK3	Loan ID

Recycled Tables Summary

Because there are similarities between the basic needs of collection management systems, especially in the areas of facilitation research and administrative and legal documentation, many tables and approaches that already exist in Specify for biological collections will be beneficial for archaeological collections as well and were therefore recycled in the database management system designed I for archaeology. However, some tables were modified and some tables were added in order to satisfy the requirements of an archaeological collection.

The aspects associated with facilitating research, i.e. being a relational database and atomization, will continue to be beneficial in a database for archaeological collections and continue in the data model for archaeological collections as they currently are. The tables and aspect associated with administrative and legal documentation will continue to be beneficial in an archaeological database, but were not entirely sufficient.

The tables for general documentation and documenting collection object condition will satisfy the needs of an archaeology collection and were left as they currently are in Specify. The current Specify tables used for tracking transactions are mostly satisfactory for an archaeological collection as they are, but a few new tables were added to satisfy unique requirements. A table for REPATRIATION was added to document a type of transaction for which archaeological collections may be accountable. PROVENANCE was also added to the data model in order to document past ownership and prove that the object was not the result of looting.

The current Specify table used to document collection object location is helpful to an archaeological collection, but is inadequate. The current Storage Location table was modified to allow the history of a collection object's locations to be documented and a table for EXHIBIT HISTORY was added to the model to better record an objects history. With these minor adjustments to the data model, research facilitation and administrative and legal documentation will be satisfied.

Revised and New Tables

With the similar database requirements taken care of, the differences between the collections must still be addressed. The two main differences discussed in the precious chapter were; the perception of collection and the role of locality. The difference in the perception of the collection requires a new approach to how to document collection object attributes since archaeological collections need to allow for many different collection object types (i.e. flutes, vessels, pins, baskets, points, etc.). The difference in the role of locality in an archaeology collection necessitates the addition of new tables in order to adequately document the locality and context from which a collection object was collected.

Perception of the Collection

The first difference between archaeological and biological collections discussed in the previous chapter is the perception of the collection. Perception was broken down into three categories; method of classification, nomenclature, and method of organization. The discussion of these categories demonstrated a vast diversity in the data requirements for archaeological collection objects. Because of the way archaeological collections are classified, named, and organized, multiple collection types are grouped together in the same collection, each with potentially different data requirements.

In order to accommodate the different classification methods and nomenclature, the table CLASSIFICATION was added to the data model. It replaces the DETERMINATION table used for biological collections. CLASSIFICATION allows the user to identify the object according to the typology and nomenclature used at their institution. TYPOLOGY in turn replaces the TAXONOMY table used for biological collections.

Figure 9 Classification Table

Classification	
PK	**Classification ID**
	Classification Date Classifier Text 1 Text 2 Number 1 Number2 Timestamp Created Timestamp Modified
FK1	Collection Object ID
FK2	Typology ID
FK3	Agent ID
FK4	Reference ID

As in the TAXONOMY table for biological collections, the TYPOLOGY table will be a tree, meaning that it will use a hierarchical relationship between the types. The tree can be defined by each institution. This allows each institution to use any typology method they prefer, whether that is one developed in-house, a more commonly used classification, such as the Getty Museum's Art and Architecture Thesaurus, or a combination of the two.

Figure 10 Typlogy Table

Typology	
PK	**Typology ID**
	Type Parent Rank Common Name Text 1 Text 2 Timestamp Created Timestamp Modified

As was discussed in the previous chapter, biological collections tend to be organized by discipline, which includes animals with similar attributes. Because of this, biological collections can use the same attribute form for their entire collection. The attributes that are used to describe one collection object are similar if not identical to the attributes that are used to describe another. For example, the attributes used to describe a hawk, such as wing span, beak length, etc. will also be used to describe a sparrow. Archaeological collections are not so uniform. They involve a variety of collection object types, each of which may require different attributes to describe it.

To accommodate this, a different approach to documenting collection object attributes was used. Table called ATTRIBUTE SET and ATTRIBUTE SET DETAIL were added to replace the COLLECTION OBJECT ATTRIBUTE table currently in the data model. The ATTRIBUTE SET table is almost identical to the COLLECTION OBJECT ATTRIBUTE table. The ATTRIBUTE SET DETAIL table is a new addition.

The difference between the ATTRIBUTE SET and COLLECTION OBJECT ATTRIBUTE tables is the way they relate to the COLLECTION OBJECT table. Currently in Specify, each collection has only one COLLECTION OBJECT ATTRIBUTE form. In the new data model, multiple ATTRIBUTE SET forms may be available.

The collection manager can decide how many different attribute forms are necessary for their collection. If it is a collection solely comprised of Roman amphorae, the collection manager may choose to only have one attribute form. However, if the collection is like most archaeological collections and is comprised of many different object types, the collection manager can choose to have many forms. If the collection has vessels, arrowheads, jewelry, swords, and beads, the collection manager can choose to have five different forms, one for each type. On the other hand, if the collection manager decides that they can use the same form for beads and jewelry, for example, he or she can decide to only have four different forms.

If there are multiple forms for the collection then once the user chooses the object type, the form created for that object will appear. In this way, the user does not have to wade through a number of irrelevant fields in order to document the relevant data for the object type in question. For example, if the user is entering data about a Roman amphora, he or she does not need to scroll past fields for documenting a sword such as shaft length, hilt length, shaft decoration, hilt decoration in order to reach the fields for the amphora. However, some attribute fields may be shared by many of the form sets. For example, attributes such as Culture, Artifact Date, Date Confidence, Technology, or Condition may be on the attribute form for each kind of object.

The ATTRIBUTE SET table includes a number of fields that may be common among different object types such as object Material, Culture, and Condition. The table also contains a number of generic fields such as "Text", "Number", or "Yes/No" to allow the collection manager to determine what attributes will be in each attribute form. These fields can be combined in any number of ways according to the desire of the curator or archeologist and according to the collection object being described. This approach allows the user to define different forms depending on the object type.

ATTRIBUTE SET has a one-to-one relationship with the COLLECTION OBJECT table. This means each collection object has one attribute description and each attribute description describes just one collection object. ATTRIBUTE SET has two one-to-many relationships with the AGENT table. These relationships are for Recorder and Measurer. This relationship means there may be more than one recorder and measurer for each collection object attribute. For example, two people might work together to document the measurements of a collection object.

Figure 11 Attribute Set Table

Attribute Set	
PK	**Attribute Set ID**
	Artifact Form
	Material
	Artifact Date
	Culture
	Condition
	Complete?
	Text 1
	Text 50
	Number 1
	Number 20
FK2	Recorder
FK1	Measurer
FK3	Artifact ID
	Timestamp Created
	Timestamp Modified
FK4	Feature ID

The ATTRIBUTE SET table has a one-to-many relationship with the ATTRIBUTE SET DETAIL table. This relationship means that one Attribute Set may have multiple Attribute Details. This is significant for objects that may have multiple parts or measurements. For example, a researcher may have a mandible with several teeth still embedded. The researcher may want to take the measurements of the mandible but also each of the teeth individually. This relationship will allow the researcher to document the basic information about the mandible such as the material, culture, and condition for the mandible once, on the ATTRIBUTE SET form. The measurements for each of the teeth can then be entered separately. This will drastically reduce the number of fields required to document relevant data, making the form much more manageable while allowing very detailed documentation.

Figure 12 Attribute Detail Table

Attribute Detail	
PK	**Attribute Detail ID**
	Artifact Type Text 1 Text 40 Number 1 Number 40 Measurer Date Measured
FK1	**Attribute Set ID**

The Attribute Set Detail table consists of the field Artifact Type to distinguish what kind of data will be documented. It also consists of generic Text and Number fields which will allow the user to customize the form to fit the needs of the artifact type. This table will function similarly to the Attribute Set table in that there will be different forms available depending on the artifact type.

The Attribute Set table also has a one-to-many relationship with the Artifact Date table. This relationship means that each collection object may have multiple dates. This relationship allows the user to assign a new date to an object if the original date is discovered to be incorrect or if a more precise date becomes available.

Artifact Date was added in order to allow the user to document the age of a collection object. Specify currently allows the user to document the age of a collection object through the date fields in the Collecting Event table and in the Stratigraphy table. For most biological collections, the date of a collection object is its death date, which is almost always the collection date, or at least very close to it. The exception is Paleontological collections.

Figure 13 Artifact Date

Artifact Date	
PK	**<u>Artifact Date ID</u>**
	Date Confidence Period Cultural Date Method Date Determiner Number 1 Number 5 Text 1 Text 5
FK1	Attribute Set ID

Similar to archaeological collections, the date of collection of a paleontological collection object has nothing to do with the object's age. Specify added the STRATIGRAPHY table to allow paleontological collections to document the age of its objects. While stratigraphy is extremely important for archaeological collections, it does not give an absolute date. It gives a relative date compared to other objects excavated from the same pit.

Instead, archaeological collections need a table that allows them to document the absolute date of an object. The ARTIFACT DATE table also has a field for Confidence, allowing the user to define how confident they are in the absolute date. The table also has fields for Period and Cultural Date, allowing the user to document another aspect of the absolute date. For example, a Mayan object might be from 300 B.C. which could also be referred to as the Late Preclassic Period.

Role of Locality

The next significant difference discussed was the role of locality in the collections. An archaeological database needs to allow for much more detail in recording the locality and provenience of an object than a biological collection. As was stated in the previous chapter, location is of utmost importance in archaeology. Location was therefore a major concern in adapting the data model for archaeological collections.

Specify currently uses three tables to document locality; LOCALITY, LOCALITY DETAIL, and GEOGRAPHY (a fourth table, GEOCOORD DETAIL, is used to document information about the locality metadata). These tables are very valuable for documenting general locality information but more tables are required in order to document locality with the specificity and relativity required of archaeological collections.

To demonstrate this need more concretely, we can use the example of an object from the collection of Dr. Dixie West, such as an Ulu knife excavated in 2006, Field Number 218. The GEOGRAPHY table would allow West to document the more general information, i.e. USA, Alaska, Aleutian Islands, Andreanof Islands group, Adak Island. The LOCALITY table would allow West to document the coordinates of the location of discovery, within a few meters of the locality, and give the location a name, such as "5 mi. SW of the intersection of Swiss Valley Road and Catfish Creek". LOCALITY DETAIL will give more information about the area in which the object was found, such as Township, Section, and Range. However, if this data are the only data recorded, a significant component of the object is lost. Also, the coordinates determined from current technology, such as a GPS handset, can be off by meters.

Archaeological localities deal with millimeters, an error range of several meters can completely change an object's significance and eliminate context entirely. Specific locality data, such as the X, Y, and Z coordinates relative to a datum, stratigraphic location, and contextual relationships are lacking. These deficiencies will be handled through the PROVENIENCE, STRATIGRAPHY, CONTEXT, and FEATURE tables. In order to add more geographic specificity, I added the table PROVENIENCE. To add more relative specificity, I added the tables STRATIGRAPHY, CONTEXT, and FEATURE.

The STRATIGRAPHY Table

STRATIGRAPHY adds another level of relative locality for objects. As mentioned in the previous chapter, paleontological collections use stratigraphy for their objects, so there is a stratigraphy table currently in the Specify data model. However, because stratigraphy is different for archaeological collections, the table for stratigraphy must be modified. Whereas paleontological stratigraphy describes strata that may extend for miles, archaeological stratigraphy may have unique characteristics for each stratigraphic element (i.e. Area, Unit, Layer, or Level).

As with the STRATIGRAPHY table currently in Specify, STRATIGRAPHY in this data model is a hierarchy. Because it is a hierarchy, a field to document which element it is added; Area, Unit, Layer, Level, etc. Although Area and Unit may not generally be considered to be elements of stratigraphy, for the purposes of this data model they are grouped together. Area and Unit are included in stratigraphy because as mentioned above, stratigraphy may be different for each Unit. Area and Unit establish in which Unit the Layer and/or Level are.

STRATIGRAPHY will have a many-to-one relationship with COLLECTING EVENT. This relationship means that each stratigraphic record may have many collecting events. For example, one particular layer may yield many collecting events.[3] Each of the collecting events can be related to the same stratigraphic element. The relationship also means each collecting event will only be attributed to one stratigraphic element. For example, in Area 1, Unit 32, Layer 5, Level B, an archaeologist may have found four pots. The pots were found far from one

[3] The collection manager can decide to treat everything collected from a layer as a single collecting event, treat each object as a separate collecting event, or treat objects excavated together as one collecting event separate from other objects or groups of objects in the same layer.

another, excavated by different archaeologists, constituting four separate collecting events. The one stratigraphic element thus has four collecting events linked to it; however, each pot will only be associated with the one stratigraphic element in which they were discovered.

Figure 14 Stratigraphy Table

Stratigraphy	
PK	**Stratigraphy ID**
	Rank Name X Y Z Condition Color Text 1 Text 2 Number 1 Number 2 **Timestamp Created** Timestamp Modified
FK1	DatumID

The STRATIGRAPHY table also has general text fields if the user wants to use different elements than Area, Unit, Layer, and Level for stratigraphy. Name was added to allow the user to document the name of the stratigraphic element. Fields X, Y, and Z were also added to allow the user to document the depth and distance north and west from a particular datum. Description fields were also added to allow the user to describe the stratigraphic element, i.e. Condition and Color.

STRATIGRAPHY also has a many-to-one relationship with PROVENIENCE. This relationship means that several stratigraphic elements may be related to a single Provenience record. This relationship was added to allow the user to document the X, Y, and Z data for the stratigraphic elements, which are taken relative to a datum. For example, the boundaries of a Unit or the depth of a Layer can be documented.

The PROVENIENCE Table

The PROVENIENCE table gives locality with a precision that no other table provides. By definition, provenience gives "the horizontal and vertical position within a matrix." (Renfrew and Bahn 2004). The PROVENIENCE table allows the user to record the relative physical position of the object at the time of discovery, offering more control of contextual information. Continuing the example mentioned above, PROVENIENCE would allow West to document the relative position of the object in the site down to the millimeter.

Unless the horizontal and vertical position is given a geographic reference, i.e. where the point or the matrix containing the point is, it is simply a point in floating space, of no use to anyone. The X, Y, Z coordinates are taken relative to a specific, known point called a Datum, from which their exact location can be ascertained. The point may be relative to an entire site or an area within the site. Many sites, if they are lucky, will have multiple objects associated with the known point.

Although called PROVENIENCE, this table includes both datum and provenience data. The two concepts are joined together because together they establish a relative locality. For database design, the two concepts can be combined because they use the same elements. By this I mean that a secondary datum and provenience both document the same data; a point with distance north or south (X), distance east or west (Y), and depth from a datum (Z). In the case of the secondary datum, the measurements are relative to the primary datum. In the case of the provenience, the measurements are relative to either the secondary or primary datum.

Figure 15 Provenience Table

Provenience	
PK	**<u>ProvenienceID</u>**
	Type (Primary, provenience) Name Elevation Latitude Longitude X Y Z Absolute X Absolute Y Absolute Z Verbatim Provenience Text1 Text2 Number1 Number2 Remarks
FK1	Collecting Event ID

A Type field was added to allow the user to document what type of record it is, i.e. Primary Datum, Secondary Datum, Provenience, etc. The Name field was added to allow the user to name the point if desired. For example, if there are multiple secondary data in a site, they may be named to differentiate them. Fields for coordinates and depth were added to document the positional data. X, Y, and Z were added to allow the user to document distances relative to a specified point. Verbatim Provenience was also added to allow the user to document less precise Proveniences. In some cases, Provenience may be a vague description, such as "Northeast corner" (Adair 2011).

PROVENIENCE has a many-to-one relationship with COLLECTING EVENT. This relationship means that each COLLECTING EVENT may have more than one Provenience. This relationship was used to allow the user to have multiple proveniences for an object or feature. For example, several provenience points may be taken for a large object or feature.

PROVENIENCE is also related to itself. This relationship means that one Provenience record may be related to another. For example, a secondary datum can be taken in reference to a primary datum; a provenience can be taken in reference to a secondary or primary datum.

Fields were also added for the absolute location at which an object was discovered. Ideally, Specify would be able to calculate the absolute location based on the distance fields and the primary datum. This would be particularly helpful in cases where a secondary or tertiary datum is used. For example, if the provenience of an object was taken from a secondary datum, Specify would be able to calculate the object location taking into account the reported distances of the object from the secondary datum and the second datum's reported distances from the primary datum.

The FEATURE Table

Whereas the other location tables are very straightforward, Feature is a unique case. The concept of Feature is murkier than provenience or stratigraphy because it is not solely a place, it is also an object. A feature is defined as "a nonportable *artifact*, not recoverable from its matrix without destroying its integrity." (Sharer and Ashmore 1979). Features can range from a bone pit which could be meters or kilometers across to a building or the wall of a building to a post hole a few centimeters wide.

Despite being defined as an artifact (a collection object), a feature by definition cannot be added to the collection. Thus, in terms of database management, a feature cannot be considered a collection object. Features are therefore accounted for in a table separate from the COLLECTION OBJECT table despite requiring much the same data.

The definition of a feature as an artifact that that is not moveable is complicated by the fact that occasionally features are moved. In Greece, entire walls have been moved to museums in order to display their frescoes (Younger 2007). In these cases, the feature is still defined as a feature but becomes a collection object as well. Thus, for the intents and purposes of database management, the feature becomes a collection object.

Objects may be collected from a feature, or part of the feature itself can be collected, but the objects collected are the objects not the feature. For example, if soil is collected from a posthole, the item collected is not the posthole; it is soil from the posthole. Further, if the constituent parts of a bone pit are collected, the bones are entered into the collection, not the bone pit. In both cases, the objects entered into the collection have ceased to be the feature except with respect to documentation of their prior relationships.

The data required for a feature is essentially the same as that of a collection object. Since it is a manmade object, it can give as much and as wide a range of information as objects do, but only if the information is recorded. Data forms used in the field for features are often similar or identical to those of objects collected (Kipfer 2007). Keeping this in mind, I linked the FEATURE table to the ATTRIBUTE SET table.

Although features may offer amazing insight, the potential insight a feature could offer is not the reason it is included in the database. The database is a collection management tool; since features are not in the collection, a feature itself does not belong in the database. A feature is represented in the database as a *locality* where collection objects may be located. Features are considered localities because this helps define where objects were excavated. Identifying the feature in which an object was found not only narrows down the exact location of the object, but gives more information about the object itself. It provides significant, additional context for the object.

Figure 16 Feature Table

Feature	
PK	**<u>Feature ID</u>**
	Associated Features **Feature Type** Feature Name Feature Number Remarks Text 1 Text 2 **Timestamp Created** Timestamp Modified

FEATURE can also define the parameters of an object's context. For example, if an object is found in what appears to be the remnants of a hole, the physical edges of that hole—indicated by the lines that record them on level plans and profiles – may be identified as representing the horizontal and vertical spatial boundaries of the object's context.

FEATURE has a one-to-many relationship with the COLLECTION OBJECT table. This relationship means more than one objects may be related to a single feature. For example, if a set of llama bones and several pots were found in a feature, each of those objects would be related to the same feature.

FEATURE has a one-to-many relationship with itself as well. This relationship means that a feature may be located within another feature. For example, a feature might be a tomb. Within the boundaries of the tomb, there may be circular pits in the floor matrix. Each circular pit is another feature. The tomb feature therefore contains several other features. The one-to-many relationship allows the user to document that the circular pits were located in the tomb feature.

In the case where objects are found in a feature and the feature is in another feature. The user would document that the objects were located in the circular pit feature. The record for the circular pit would document that the pit, and therefore the llama bones and pots, was located in the tomb feature. The llama bones and pots would be related to the tomb through the pit.

Similar to the COLLECTION OBJECT table, the FEATURE table has a many-to-one relationship with COLLECTING EVENT and has a one-to-one relationship with the ATTRIBUTE SET table. The many-to-one relationship with the FEATURE table means that a collecting event may be related to more than one feature. This allows the collection manager to define a COLLECTING EVENT as broadly as desired. COLLECTING EVENT can either be used for each Feature or Object excavated, or it can be used for a broader array of excavation activities.

The one-to-many relationship with the ATTRIBUTE SET table means that each feature may have one attribute record associated with it. The attribute record will allow the user to document the different characteristics of the feature similar to the way in which a collection object's attributes would be documented. The ATTRIBUTE SET table will allow the user to document aspects such as Feature type, i.e. bone pit, differing matrix, wall, etc., dimensions of the feature, and other details as appropriate. Similar to the relationship between COLLECTION OBJECT and ATTRIBUTE SET, multiple attribute forms will be available for features depending on the feature type. As in the COLLECTION OBJECT ATTRIBUTE SET the number of forms available will depend on the collection manager.

The CONTEXT Table

Context, as has been noted by so many archaeologists and several times in this paper, is everything. Context is defined as "Characteristics of archaeological data that result from combined behavioral and transformational processes, evaluated by means of recorded association, matrix, and provenience" (Sharer and Ashmore 1979). As noted, context offers information on a variety of aspects of the culture that produced the remains but the information is gained through objects' spatial relationships with other objects. It is considered "essential documentation for artifacts" and has been considered as such since the nineteenth century (Krakker *et al*. 1999).

Surprisingly, as important as context is for an object, most databases do not include a means of documenting context and certainly not with any specificity. A table was added to this model in order to allow the user to document context. Context can be as widely or narrowly defined as desired by the archaeologist. Where the boundaries for one context are drawn is decided individually, whether it be a posthole, an entire house, or a complex of houses. In the model I did not attempt to limit the definition of context. I allow users to define it as liberally or conservatively as required for their study and data.

Figure 17 Context Table

Context	
PK	**<u>Context ID</u>**
FK1	**Related Artifact** Relationship Type Text 1 Text 2 Number 1 Number 2 Remarks **Timestamp Created** Timestamp Modified

Relationship Type was added to allow the user to document the exact nature of the relationship between objects. For example, the relationship between a metal plate found in the general proximity of human remains has a different context from a metal plate that is found in the mouth of a mummified human head. The Spatial Relationship field allows the user to say if the related object was found in, on, under, or next to another object.

The CONTEXT table is directly linked to the COLLECTION OBJECT table through a many-to-one relationship. The many-to-one relationship means that one collection object may have multiple contextual relationships with other collection objects. In other words, a body excavated in a tomb can be contextually related to the metal plate found in an individual's mouth, the beads found on the sternum of the skeleton, the spindle whirls found in the skeleton's left hand, the obsidian blade found in the skeleton's right hand, the ceramic pot found between the ribs and the right arm, and the *Spondylus* shell found under the head.

CONTEXT also has a one-to-many relationship with itself. This relationship means that the user can document contexts within contexts. For example, in a tomb, there may be a pedestal with a body and various objects. The archaeologist may decide that the pedestal is one context within the larger context of the tomb. Further, the tomb may be in an area with a series of tombs, which the archaeologist may consider a context. In this example, there are three contexts, each within the other.

Specify programmers have recently instituted a concept called "Container" into the Specify data model. The intention of the Container table is to allow multiple collection objects to be related when they share a physical or intellectual container. In other words, the objects would be in a container if they are in the same jar (a physical container) or if they were discovered in the same stomach contents (intellectual container). The Container table allows the container itself to be documented as well as the objects in or on it. For example, in botanical collections, a sheet may have multiple leaves on it. It is not unknown for botanists to document both the sheet and the leaves on the sheet (Beach 2008). However, they want to maintain the relationship of the objects to the sheet and to each other through the catalog number and container allows them to do so.

Container functions in the same way CONTEXT was intended to function in my data model. Similar to CONTEXT, Container allows the user to document an intellectual relationship (i.e. not being physically kept together) between different objects. However, as the Container table functions now, it does not allow for specific documentation of the relationship. Additional fields, such as Spatial Relationship, will have to be added in order to adequately document an object's context.

Revised and New Table Summary

Because archaeological collections are perceived of differently than biological collections, several tables had to be added or modified in order to accommodate archaeological data. TYPOLOGY and CLASSIFICATION are modifications of the current Determination and Taxonomy tables in Specify. Because an archaeological collection may include many different types of objects, ATTRIBUTE SET was added to allow the user to document attributes of an object without having to scroll through fields for other types of objects. ARTIFACT DATE was added to allow the user to document an object's original date rather than just the date the object was collected.

To satisfy the Role of Locality difference, I added the STRATIGRAPHY, PROVENIENCE, FEATURE, and CONTEXT tables. These four tables will allow the archaeological user to document the locality of the collection objects with a specificity that is required of an archaeological collection but was not possible with the previous Specify locality tables. It is possible that the recent work adding the Container field may help resolve many of the same issues that CONTEXT was added to resolve. In this case, the fields from the proposed CONTEXT table can be added to the Container table. This would fulfill the archaeological collection need to document context without having to add a new table.

Conclusion

There are many similarities between an archaeological and a biological database requirements. These allow for use of many approaches and tables currently in Specify's data model. Specify's relational database and atomization approaches were continued in the data model for archaeological collections. Tables to address the need for general documentation and condition documentation did not require modification because they satisfy the needs of an archaeological collection as they are.

The tables to track collection management transactions were also almost completely satisfactory as they were. The only need left unfulfilled was the need to document Repatriation, so a new table was added. The tables for object location needed some modification to satisfy archaeological needs, so the table Storage Location was modified to allow for multiple locations, both past and present. A table for the storage container was added to make locating a collection object quicker. Finally, a table for EXHIBIT HISTORY was added to more fully document an object's past locations and exposure to other objects.

However there are significant differences between the needs of these two collections as well. The two main differences discussed were related to the perception of a collection and the role of locality. These two differences necessitated the addition of several tables to the data model and the modification of several others in order to satisfy the requirements of an archaeological collection.

The first main difference is the perception of the collection. The current Determination and Taxonomy tables were only slightly modified to accommodate archaeological collections, becoming CLASSIFICATION and TYPOLOGY. Because archaeological collections may be comprised of different object types, with different attributes to be recorded, a new approach to documenting object attributes was implemented. ATTRIBUTE SET was added in order to document object and feature attributes. This table will offer different attribute forms depending on the type of object or feature described. This allows the user to enter data quickly and efficiently, without having to scroll through more fields than necessary. Because archaeological collections involve objects that are older than the date they were collected, ARTIFACT DATE was added to the database. This allows the user to document the date of deposition rather than just the date collected.

Role of locality was the other significant difference. STRATIGRAPHY was added to allow the user to document the strata in which an object or feature was discovered. Although STRATIGRAPHY is currently in Specify, stratigraphy as used by archaeology is different. Therefore, the STRATIGRAPHY table is completely new to the data model. PROVENIENCE was added to allow the user to document the relative position of an object, potentially down to the millimeter. This offers a specificity that is not possible in Specify currently. FEATURE was added to allow the user to document where an object was discovered with more specificity. FEATURES describe the locality at which an object was discovered and can offer much more insight into an object. Finally, CONTEXT was added to allow the user to document the context in which an object was discovered.

Chapter 6 – Conclusion

Throughout this thesis and my presentation of the data model, I have attempted to create a guide that programmers can use to adapt Specify for archaeological collections. While software such as Open Context or tDAR allows archaeologists to share their data, there are no free, open-source collection database management systems designed specifically for archaeological collections. The CDM systems currently used by allow are not as effective as they could be. The data model I propose addresses archaeology-specific issues that have not been effectively resolved by other archaeological collection database management software. Current archaeological collection DBMS has been adapted from that used for other types of collections without adequately resolving significant requirements for archaeological collections such as variety of object types and the need to document specific and relative locality information such as CONTEXT and PROVENIENCE.

Although archaeological collections have many different attribute requirements, the available DBMS's treat archaeology as a unified collection with only one attribute form for item description. KE EMu, The Museum System (TMS), PastPerfect, FileMaker Pro, and Microsoft Access are among the more popular systems. The Yale Peabody Museum uses KE EMu (Teasley 2011). The Field Museum anthropology collection was using FileMaker Pro but in 2008 decided to switch to KE EMu (Tilson 2008). The Sam Noble Oklahoma Museum of Natural History (SNOMNH) archaeology collection uses a system they created in house (Leith 2008). The J Paul Getty Museum uses TMS (TMS website 2011).

However, systems such as TMS, PastPerfect, Re:Discovery, and KE EMu, use the same forms for all objects in the collection. KE EMu even states that they differentiate forms based on discipline so "users cataloging pottery do not see fish fields" (KE EMu website 2011) but make no differentiation between pottery, projectile points, or soil samples. Since they are categorized by the same discipline, they have to use the same fields. In fact, the only DBMS I found that offers different forms depending on the object type is the one created at the Sam Noble Oklahoma Museum of Natural History. Context and provenience are of supreme importance in archaeology. However, these have been ineffectively modeled and represented in current commercial DBMSs.

In order to effectively document an archaeological collection, these concepts must be taken into account not only for data entry but for data retrieval. Several of the available DBMSs offer ways to associate objects, providing some context, but they ignore features, a significant part of context for many objects. For example, KE EMu offers only a single field to describe FEATURE. FEATURE is treated as a location with no more description of the feature entered. This means the object aspect of FEATURE is not recorded. PROVENIENCE in one form or another is included in most DBMSs but it cannot be relative. In systems such as PastPerfect, TMS, and KE EMu, collections can document the latitude, longitude, and depth of an object, but these are absolute values. As mentioned in Chapter 3, archaeologists record these values relative to a datum. However, the DBMSs do not offer the ability to document the datum and relate PROVENIENCE to it.

Specify is freeware and open-source, meaning that not only is the software free for anyone to install and use, but the code used to create it is freely available to anyone interested in reviewing or modifying it. It also uses only other freeware programs such as Java and MySQL. Specify is actively supported by programmers who are constantly updating it to add new features and make it more efficient and effective based on user feedback and requests. It is an intuitive program intended to make data entry and retrieval as efficient as possible. For those with experience with relational databases, no training is required to use Specify effectively. For those who would like a more in depth understanding of Specify and its features or for those who are not comfortable with relational databases, there is a user manual describing almost all aspects of the program and training sessions are offered. Interested parties can either attend one of the Specify workshops held once a year at the University of Kansas or have a Specify representative come to their institution to demonstrate how to use Specify and how to take advantage of its many features.

Support and customization of Specify is free for US non-profit institutions, allowing the users to make a number of modifications to the database to accommodate their needs. It includes a web interface, allowing the users to put their data online, or at least the parts of the data that the user wants to make accessible to the public. These aspects of Specify make it available to any museum or academic research center with a computer and Internet access regardless of budget. That means that even museums with small budgets can adequately and efficiently document their collections and make the data available to the public.

Data conversion in Specify is also free. Specify offers two ways of transferring data from a current system into Specify. First, Specify staff can transfer data for other institutions, importing it from any electronic system into the Specify database. Second, a Specify feature called the WorkBench allows a user to upload data him- or herself. This spreadsheet-based application allows the user to enter data in a spreadsheet format or import data from an external spreadsheet. Once data are in the WorkBench, a user can "map" it, or say what field in Specify a spreadsheet column should enter. For example, the user can indicated that the column labeled "Date" should go into the Specify field called Collection Date.

This thesis evaluated whether Specify could be adapted to be as effective for archaeological collections as it is for biological collections. The first step was to investigate the methods used by archaeological collections and the data documented in the field and once the collected objects are in the collection. The next was to conduct a comparison of archaeological and biological collections database management needs. The comparison indicated what changes needed to be made to the Specify data model in order to accommodate archaeological collections. It identified many similarities between the two collections, including the need to effectively enter, search, and share data, and record administrative information. It also identified some differences between the two databases, including differences in classification method, object nomenclature, organizational method, and the specificity required for documenting locality.

My analysis demonstrates that Specify can be adapted to effectively document archaeological collections and allow for efficient data retrieval. Many of the tables already in the Specify data model will be useful for an archaeological collection with little or no modification. Tables that will have to be added though in order to accommodate the specific needs of an archaeological database those for documenting specific locality and relative locality and those for allowing for different types of collection objects to be documented, each type with different data requirements.

Ideally, this data model would make every datum an archaeologist could want readily available and easily discovered by researchers. However, the data model will fall short of the ideal for several reasons. First, it is impossible to anticipate every datum that might be desired. Second, if every datum was recorded in a different field, data entry and search would be clumsy to the point of being more of a burden than an advantage.

I attempted to compensate for the first issue through the use of a number of generic fields that can be used to document data as the user sees fit and to compensate for the second issue by offering different forms depending on the object type. This will cut down on the number of fields on the data entry form, ideally culling out the fields that may be necessary for one object type, but not another.

The data model falls short of the ideal because it does not attempt to create, nor does it use a common classification terminology. However, this allows each institution to use and enforce whatever system they want to use, whether that be a commonly used one, such as the Getty Museum's Art and Architecture Thesaurus, or an institutional one such as is used at the Kansas State Historical Society. This will facilitate data entry and searching within a given institution by controlling vocabulary. However, without cross-institution standards, searches between databases of different institutions will continue to be problematic. Using a particular search term to search one collection may not give results in another collection even if they have similar objects.

I have attempted to anticipate differing needs by drawing from my archaeological experience as well as interviewing people from different institutions and archaeologists working in different areas of the world. I have considered records for different object types with differing data and I have used the data model to document real objects. However, I cannot be sure if I adequately addressed the needs of archaeological collections until the database system is created and used. Actual use of the database may reveal things that were missed in my data model.

Adapting the data model is just the first step in the process of creating DBMS software. A data model establishes how data should be organized and related. It says nothing about how to implement the data model nor does it make any program architecture suggestion. Based on the data model, a database schema will be created, which will be a more inclusive and elaborate description of the model. The database schema includes behind-the-scenes tables that are necessary from a programming perspective to enable the relationships described in the database model. For example, the database model may say that there is a field for Catalog Number but the Database Schema will include a table for the Catalog Numbering Format, which the user will never see but from which he or she will benefit.

From the schema, software programmers can create a working model of the DBMS, which will hopefully be dispersed to museums around the world for their archaeological collections. The programmers may create a Beta version, which will be tested by archaeological collections. Based on the results of those tests, adjustments will be made, features may be added and relationships may be changed as users sample the software and make suggestions to make it more effective.

Once a working model is created, museums and academic collections, whether large or small, will be able to proficiently document their collections and make them accessible. Hopefully, this program will be particularly appealing to smaller institutions that may not have much funding but want to be able to effectively document their collection and make it available to the public. It will also be appealing to institutions that are already using Specify for their biological collections and would like to use the same system for all of their collections. In fact, several archaeological collections are already using Specify (unsupported by Specify) for this exact reason, including the College of Idaho and the Cranbrook Institute of Science. Several other institutions, such as the University of Wisconsin, Madison, Louisiana State University, Baton Rouge, and the Bishop Museum in Honolulu have also showed interest in using Specify for all of their collections, including their archaeological collections, if Specify can be modified to accommodate archaeological collections.

The data model I have proposed will allow for more effective documentation of provenience and collection object attributes. Provenience can also be more effectively documented because it can be documented as it was recorded (as a relative measure instead of an absolute measure). With the technology available today and the plugins used by Specify, it is likely Specify will be able to calculate the relative provenience from the datum and graphically show the positions of the objects, enabling users to see context in a way they could not in other DBMS's.

The more effective documentation of collection object attributes will allow more effective data retrieval of the data as well. Because different forms are used for different object types, users can enter data in the fields relevant to the object they are describing. The effect is that more data can be documented in separate fields, which enables more effective searching and exporting of relevant data for analysis.

Collections with data that have been previously inaccessible can be accessed. Specify will allow archaeologists and researchers to search the collections of remote or small museums or academic collections that they may never have been able to search before. Small museums may only contain a small part of an area's history. In-depth analysis of objects in the collection can be done but comparative analyses are considerably more difficult.

Increased accessibility and more comprehensive records allow for investigations such as tracing migration patterns, tracing cultural influence, or the introduction of tools. The level of access and collaboration allows investigators to ask questions of archaeological data that would otherwise not be technically approachable. It also allows for data integration to create merged datasets that can give a broader picture of the development of a region both culturally and biologically and the interaction of the two.

The advantages of the proposed information model are: 1) increased accessibility, 2) interdisciplinarity, and 3) capacity for effective record keeping. The limitations of the proposed information model are: 1) the fact that the model has not been tested with actual database information, 2) advances in technology that will occur in information representation, database architectures, hardware platforms, cannot be fully anticipated, and 3) software programmers have not yet attempted to use the model to create software. Because of this, the data concept relationships described in the information model mayor may not be practical.

Specify is currently used by over 247 institutions around the world. However, it was not designed for nor does it yet support archaeological collections. Even without adaptation, many of the features that Specify offers natural history collections would be advantageous for archaeological collections. With some adaptations, Specify can be as effective and efficient at documenting archaeological collections as it is for natural history collections.

Bibliography

Adair, Mary. Curator, Archaeological Research Center, University of Kansas. Interview October 6, 2007.

Adair, Mary. Curator, Archaeological Research Center, University of Kansas. Interview September 30, 2011.

Adams, William Y. Ceramic Industries of Medieval Nubia. Memoirs of the Unesco Archaeological Survey of Sudanese Nubia, vol. 1, 1986.

Adams, William Y. and Ernest W. Adams. Archaeological Typology and Practical Reality: A Dialectical Approach to Artifact Classification and Sorting. Cambridge: Cambridge University Press, 1991.

American Association of Museums. AAM Board of Directors approved "The Accreditation Commission's Expectations Regarding Collections Stewardship" December 17, 2004.

American Association of Museums. AAM Board of Directors approved "AAM Standards Regarding Facilities and Risk Management" (http://aam-us.org/resources/ethics-standards-and-best-practices/characteristics-of-excellence-for-u-s-museums/facilities-and-risk-management), July 2007.

Arp, Ryan S. "An ArcGIS Data Model for Archaeologists: Problems and Prospects" 24th Annual ESRI International User Conference, 2004.

Banning, E.B. The Archaeologist's Laboratory: The Analysis of Archaeological Data. New York: Kluwer Academic, 2000.

Beach, James. Director, Specify Software Program, Biodiversity Research Center, University of Kansas. Interview January 12, 2008.

Bentley, Andrew. Ichthyology Collection Manager/Specify Usability Lead, Biodiversity Research Institute, University of Kansas. Interview February 18, 2008.

Burger, Richard L. and Lucy C. Salazar. Machu Picchu: Unveiling the Mystery of the Incas. New Haven: Yale University Press, 2004.

Carmichael, David, Robert Lafferty III, and Brian Leigh Molyneaux. Excavation Archaeologist's Toolkit Volume 3. New York: Altamira Press, 2003.

Carr, Christopher. ed. For Concordance in Archaeological Analysis: Bridging Data Structure, Quantitative Technique, and Theory. Kansas City: Westport Publishers, Inc., 1985.

Castillo Butters, Luis Jaime. Scientific Director San Jose de Moro Project. Interview July 3, 2008.

Chenhall, Robert. Nomenclature for Museum Cataloging: A System for Classifying Man-Made Objects. Nashville: American Association for State and Local History, 1978.

Chenhall, Robert G. "Computerized Data Bank Management" Data Bank Applications in Archaeology. Sylvia Gaines ed. Tucson: University of Arizona Press, 1981.

Codd, E. "A Relational Model of Data for Large Stored Databanks". Communications of the ACM. v. 13 (6), June 1970.

Coggins, Kathy. Project Manager, Specify Software ProjectBiodiversity Research Institute, University of Kansas. Interview March 18, 2008.

Connolly, Thomas M. and Carolyn E. Begg. Database Systems: A Practical Approach to Design, Implementation, and Management. Reading, MA: Addison-Wesley Longman, Inc, 2002.

Corbett, Debbie. "Question" Email to Dixie West. May 19, 2009.

Cowgill, George. "An Update on Teotihuacan" Antiquity. (82) 2008.

Darrow, Barbara. "Whatever Happened to Lotus 1-2-3?" CRN February 1, 2002.

Denning, Kathryn. "'The Storm of Progress' and Archaeology for an Online Public". Internet Archaeology. 15 (15), January 2008.

Dibble, Harold and Shannon McPherron. "On the Computerization of Archaeological Projects." Journal of Field Archaeology. 15, 1988.

Farrell, Una. Collections Manager Invertebrate Paleontology, Biodiversity Research Center, University of Kansas. Interview September 30, 2011.

Gaffney, V. and S. Exon. "From Order to Chaos: Publication, Synthesis and the Dissemination of Data in a Digital Age." Internet Archaeology. 6 (6), 1999.

Gaines, Sylvia. "The Impact of Computerized Systems on American Archaeology: An Overview of the Past Decade." in Information Systems in Archaeology. ed. Roger Martlew. Gloucester: Alan Sutton Publishing, 1984.

Gardin, Jean-Claude. "Archaeological Discourse, Conceptual Modeling and Digitalization: An Interim Report of the Logicist Program." Report to the CAA, 2002.

Garneau, Glenn, Data Manager, Specify Software Program, Biodiversity Research Center, University of Kansas. February 4, 2008.

Garst, Christine. Archaeology Lab, Kansas State Historical Society. Interview September 14, 2007.

Getty Art and Architecture Thesaurus Online, The Getty Research Institute, J. Paul Getty Trust, 2004.

Glassow, Michael. "Excavation" Handbook of Archaeological Methods. H. Maschner and C. Chippendale eds. Lanham: AltaMira Press, 2005.

Green, Chris. "Ary Sentence Hinges on Value of Loss" The Hutchinson News. May 12, 2005.

Green, Chris. "Jury Convicts Ary on 12 Counts" The Hutchinson News. November 2, 2005.

Harris, Edward. Principles of Archaeological Stratigraphy. London: Academic, 1989.

Hoffman, C. Marshall. "Projectile Point Maintenance and Typology: Assessment with Factor Analysis and Canonical Correlation" For Concordance in Archaeological Analysis: Bridging Data Structure, Quantitative Technique, and Theory. Christopher Carr ed. Kansas City: Westport Publishers, Inc., 1985.

Innes, Megan. Intern for the Collections and Archives in the Anthropology Division, American Museum of Natural History. Interview June 30, 2008.

Kansa, Eric. "Report on the Digital Data Interest Group (DDIG): Presented to the Society for American Anthropology Executive Staff and Board of Directors". February 2007.

Kansa, Eric. "A Community Approach to Data Integration: Authorship and Building Meaningful Links Across Diverse Archaeological Data Sets." 2007

KE EMu Electronic Museum. 2011. KE Software website. October 29, 2011. (http://www.kesoftware.com/emu)

Kintigh, Keith. "The Promise and Challenge of Archaeological Integration." American Antiquity. 71 (3), 2006.

Kipfer, Barbara Ann. The Archaeologist's Fieldwork Companion. Oxford: Blackwell Publishing, 2007.

Krakker, James, David Rosenthal, and Deborah Hull-Walski. "Managing a Scholarly Resource: Archaeological Collections at the National Museum of Natural History". Museum Anthropology. 23 (2), 1999.

Kvamme, Kenneth. "Terrestrial Remote Sensing in Archaeology" Handbook of Archaeological Methods. H. Maschner and C. Chippendale, eds. Lanham: AltaMira Press, 2005.

Leith, Liz. Archaeology Collections Manager, Sam Noble Oklahoma Museum of Natural History. Interview January 10, 2008.

Liu, Sofia Galarza. Collection Manager, Spencer Museum of Art. Lecture April 2, 2008.

Lock, Gary. Using Computers in Archaeology: Towards Virtual Pasts. New York: Routledge, 2003.

Lockhart, Jami. Discussion Session, Archaeoinformatics Workshop, Biodiversity Research Institute, University of Kansas, June 6-7, 2007.

Martlew, Roger. Information Systems in Archaeology. Gloucester: Alan Sutton Publishing, 1984.

Maschner, Herbert and Christopher Chippendale ed. Handbook of Archaeological Methods. Lanham: AltaMira Press, 2005.

McAnany, Patricia A. ed. K'axob: Ritual, Work, and Family in an Ancient Maya Village. Los Angeles: Cotsen Institute of Archaeology University of California, 2004.

McCorry, Helen and Ian O. Morrison. "Report of the Catechism Project". National Museums of Scotland, 2005.

McPherron, Shannon and Harold Dibble. ed. Using Computers in Archaeology. Mayfield: McGraw-Hill, 2002.

Miller, John. Museum Specialist, Collections Processing Unit. National Air and Space Museum, Garber Center. Interview October 14, 2008.

Miller, John. Museum Specialist, Collections Processing Unit. National Air and Space Museum, Garber Center. Interview August 13, 2008.

Morris, Paul J. "Relational Database Design and Implementation for Biodiversity Informatics." PhyloInformatics. 7, 2005.

Odess, Dan Discussion Session, Archaeoinformatics Workshop, Biodiversity Research Institute, University of Kansas, June 6-7, 2007.

Olien, Michael. "Information Management Programs for Anthropologists." Culture & Agriculture. 12:23-26, 1992.

Ollenburger, Jeff. “Kansas Cosmosphere and Space Center Official Statement Regarding Charges Against Former Cosmosphere President Max Ary”. SpaceRef. April 8, 2005.

PastPerfect Software. 2009. PastPerfect Software, Inc. September 30, 2009 (http://www.museumsoftware.com/)

Pearlman, Robert. “Former Cosmosphere Director Indicted” CollectSPACE. April 7, 2005.

Planas, Dawn. Curatorial Assistant, Museum of the North, University of Alaska. Interview February 28, 2008.

Purdy, Barbara A. How to do Archaeology the Right Way. Gainesville: University Press of Florida,1996.

Reilly, Paul and Sebastian Rahtz. ed. Archaeology and the Information Age: A Global Perspective. New York: Routledge, 1992.

Renfrew, Colin and Paul Bahn. Archaeology: Theories, Methods, and Practice. London: Thames and Hudson, 2004.

Richards, Julian and Damian Robinson. ed. Digital Archives from Excavation and Fieldwork: A Guide to Good Practice. Oxford: Oxbow Books, 2000.

Richards, J.D. and N.S. Ryan. Data Processing in Archaeology. Cambridge: Cambridge University Press, 1985.

Richardson, Steve. “Followup from Re:discovery Software” Email to Theresa Miller. September 9, 2009.

Rickard, T.A. “The Nomenclature of Archaeology.” American Journal of Archaeology. 48 (1), 1944.

Ryan, Nick. “Databases.” Internet Archaeology. 15 (15), 2004.

Sarasan, Lenore. “Why Museum Computer Projects Fail”. Museum News. Jan/Feb 1981.

Shannon, Kerrie-Ann. Curatorial Assistant, Museum of the North, University of Alaska. Interview February 28, 2008.

Sharer, Robert and Wendy Ashmore. Fundamentals of Archaeology. Menlo Park, California: The Bengamin/ Cummings Publishing Company, Inc., 1979.

Snow, Dean et al. “Cybertools and Archaeology”. Science. 113, February 17, 2006.

Stewart, R. Michael. Archaeology: Basic Field Methods. Dubuque, Iowa: Kendall/Hunt Publishing, 2002.

Sullivan, Lynne and S. Terry Childs. Curating Archaeological Collections: From the Field to the Repository. Archaeologist's Toolkit Volume 6. New York: Altamira Press, 2003.

Teasley, Stephanie. Collections Assistant, Spencer Art Museum, University of Kansas. Interview March 16, 2011.

Tilson, Misty. Registrar, Department of Anthropology, The Field Museum. Interview January 4, 2008.

West, Dixie. Adjunct Research Associate in Archaeology, Kansas State University. Interview February 13, 2008.

Whitney, James. Collection Manager, Museum of the North, University of Alaska. Interview February 28, 2008.

Wisniewski, Josh. Curatorial Assistant, Museum of the North, University of Alaska. Interview February 28, 2008.

Younger, John. Curator, Wilcox Classical Museum, University of Kansas. Lecture, October 16, 2007.

Yeung, Albert and G. Brent Hall. Spatial Database Systems: Design, Implementation, and Project Management. Dordrecht: Springer, 2007.

Appendix 1 – Glossary

Agent – Any person, group of people, or institution that does something related to an object in the collection. For example, an agent may be an excavator, curator, cataloger, photographer, accessioner, etc. An agent may also be a donor or sponsor which may be an individual or an institution such as the National Science Foundation.

Atomization – Entering each piece of data into a separate **field**.

Attribute – A physical characteristic of an object such as form, medium, height.

Attribute Set – A set of **attributes** pertaining to an object.

Collecting Event – The collecting of objects as defined by a specific day, person or people, location, collection methods, and possibly time.

Collection Object – An object intended to be entered into the collection, usually marked with a catalog number.

Context – The area around an artifact when excavated, including its matrix and geographically related artifacts. This may be defined as narrowly or broadly as the excavator decides.

Data Model – A representation of how data concepts are related to other data concepts in a database.

Database – A "shared collection of logically related data, and a description of this data, designed to meet the information needs of an organization" (Connolly and Begg 2002). See Appendix 2 for further description.

Database Management System (DBMS) – A computer program that allows one to store, edit, organize, and search data.

Datum – A point with coordinates and an elevation used as a reference for other data or proveniences.

Debitage – Artifacts representing waste produced while creating tools of stone, bone, or other material.

Feature – An artifact that cannot be moved, such as a bone pit, post hole, or a wall.

Field – A space in a database for holding a piece of digital data. See Appendix 2 for further description.

Foreign Key – A code used by one table in a **relational database** to link that table to another table. It matches the **primary key** of the linked table.

Form – The recording method or user interface, such as a paper form or a data entry display on a computer screen, used to enter data that are part of a **record**.

Geographic Information System (GIS) – "Software-based systems designed for the collection, organizing, storage, retrieval, analysis, and displaying of spatial/digital geographical data held in different 'layers'." (Renfrew and Bahn 2004)

Locality – The Location where an object was collected.

Many-to-Many Relationship – A **table** may have one or more records related to one or more records from another table. For example, a **collection object** may be related to more than one **agent** (excavator, recorder, photographer, etc.) and an agent may be related to more than one collection object.

Matrix – The physical environment in which an artifact is discovered.

One-to-Many Relationship – A **table** may have a record related to more than one record from another table. For example, a **collecting event** may be related to more than one **collection object** if multiple objects were collected at the same time. However, a collection object is only collected once, so it will only have one **collecting event**.

One-to-One Relationship – A **table** will have one record related to one record from another table. For example, a Collection Object will be related to one Attribute Set record. That Attribute Set record will only be related to a single Collection Object.

Primary Datum – The main datum for a site which other data or proveniences are taken.

Primary Key – The code used in a relational database to identify a table. See Appendix 2 for further description.

Provenance – The ownership history of an object (i.e. who has owned the object).

Provenience – An artifact's horizontal and vertical position in a **matrix**. Often taken as a distance north or south, east or west, and up or down from a known **datum**.

Record – A set of "logically related data" (Connolly and Begg 2002)

Relational Database – Database in which tables are linked together through at least one common linking field. See Appendix 2 for further description.

Schema (including **database schema**) – The design of a database.

Secondary Datum – Datum with coordinates and elevation taken in relation to the primary datum. Used as a reference for other data or proveniences.

Site – An *in situ* spatial clustering of artifacts, features, and other evidence of human activity (Renfew and Bahn 2004). See Appendix 2 for further description.

Table – A set of existing fields in a relational database that have been organized in a specific way. A table is conceptually organized as horizontal rows and vertical columns.

Type – An idealized, hypothetical artifact used to characterize a group of artifacts that share a common and recognizable set of attributes.

Typology – A systematic collection of types used to organize artifacts, features, sites, or other archaeological phenomena.

Appendix 2 – Database Terminology

The first and perhaps the most basic term that must be understood is *database*. According to the definition I will be using here, a database is a “shared collection of logically related data, and a description of this data, designed to meet the information needs of an organization” (Connolly and Begg 2002). Essentially, it is a way to keep track of information, as in the case of museums, about collections. Databases may occur in a variety of formats, including paper archives in bound record books or cardfiles as well as digitized, machine-readable computer files.

At present, the most useful format for large databases in museums is in the form of digital files that can be read and displayed via the computer. Museums such as the Field Museum in Chicago, the National Air and Space Museum (NASM) in Washington, D.C., the Sam Noble Oklahoma Museum of Natural History (SNOMNH) in Normal, Oklahoma, and the Museum of the North in Fairbanks, Alaska, all use software systems to database their collections. The systems used range from Microsoft Office Access to EMu to The Museum System to one created in house (SNOMNH).

A *relational database* documents data through the use of interrelated sets of information, called tables, which reference each other. Data are broken up into variables, recorded as fields in different tables. The tables are through the use of common *fields*, i.e. fields such as a primary key ID that is duplicated in the linked tables (McPherron and Dibble 2002). A primary key ID is “a field or set of fields in an index used to find and sort records in a database table” (McPherron and Dibble 2002). In other words, the primary key is what links one table to another table containing associated information. The primary key ID is unique and arbitrary; usually the users do not see it. It serves purely for the purpose of linking data tables.

For example, information on the loan of an object may be separated into a Loan table and an Agent table. The Loan table may contain data about the loan such as what items were loaned, when they were loaned, when they are due back, and who approved the loan. The Agent table may contain a person's first, middle, and last name, their position, title, email, and an agent ID (the primary key ID). The Agent ID may be assigned automatically and may not be seen by the user. If John Smith approved loan 001, John Smith's agent record would contain his first and last name, any other data the user has for him, and a unique Agent ID. The loan John Smith approved would contain the loan number, other related data, and John Smith's Agent ID (primary key ID). When the user views the record for Loan 001, they will see the loan data entered and be able to see John Smith's data because the Agent ID is in the Loan record.

A *table* is a set of existing fields that have been organized in a specific way. It is made up of rows and columns. The table contains variables, known as fields, which represent diverse kinds of information such as, people involved with the artifact (i.e. agents, excavators, recorders, photographers, treatment agents, etc.), loans, treatments, attributes, and geography, among others. The columns are the fields in tables and the rows are the records.

A *field* is the name of a variable that records a specific piece of data. It can appear visually as a space on a table where data are entered (that is, it can be thought of as the blank lines on a paper form where information is intended to be written). The same information that can be written in an open space on a paper form can be entered into a database field. There are limits on the amount of data that can be entered into each field, which can be set differently for each field. Fields can document data in different formats including nominal, presence/absence, numeric, or text. In some programs, the values entered into the fields can be controlled by format or options available, i.e. allowing only a "day, month, year" format versus "month, day, year" format in a date field or by choosing only among ADK-011, ADK-013, etc. in the field for archaeological site number.

Once the data are entered into a table and saved, it becomes a record. A *record* is a set of "logically related data" (Connolly and Begg 2002). A record is distinct from a table in that a record is one row of data in a table. It consists of data from the fields in the table. However, there may be more information included in the record than can be contained in that one table. The record may also be comprised of variables recorded on several tables if artifact data include other information, such as information on the site where the artifact was collected (recorded on the Locality table), information on who excavated the artifact (recorded on the Agent table), and information on the physical properties of the artifact (recorded in the Attribute Set table), to name a few. A record can therefore be comprised of data recorded from fields represented on a variety of different tables all linked through the primary key ID.

A *form* is a way to enter and view data. It is a user interface for data in a table. A table is different than a form. Data are entered into a table through a form. Whereas a table is a list of fields and the data in the fields, a form is purely where the data are entered or viewed. It does not contain data. A form is essentially a digital template, similar to a paper form that might be filled out to document a collection object.

Appendix 3 – Database Management Systems Discussed

- 4D, 4D SAS (www.4d.com)
    - Pros
        - Highly user friendly
        - Multi-platform operable
        - Can use sub-numbers
    - Cons
        - Cannot attach photographs to collection objects
        - Not all of the fields are searchable
        - No field for excavation technique
        - Cannot document who last edited a record
        - Cannot track changes to the Collection Object record
- FileMaker Pro, FileMaker, Inc. (www.filemaker.com)
    - Pros
        - Allows for barcodes
        - Can use sub-numbers
        - Can document Exhibit History
        - User customizable
        - Offers Customer Support
    - Cons
        - Cannot be shared my multiple departments
        - Limited ability to attach files
        - Cannot print customized reports or labels
        - High cost for a software license

- The Museum System (TMS), Gallery Systems (www.gallerysystems.com/tms)
    - Pros
        - Can attach files to various records
        - Can carry-forward information so you do not have to re-enter data if records have the same information
        - Can update location of objects in a batch so if you are moving multiple objects, you can update all of them at the same time
    - Cons
        - Could not track past storage locations
        - Could not document conservation procedures
        - Cannot associate collection objects with other objects
        - Cannot document precise provenience
- Microsoft Office Access 2003, Microsoft Corporation (office.microsoft.com/en-us/access)
    - Pros
        - Comes with the Microsoft Office Suite, which most people have already so it may seem free
        - Highly customizable
    - Cons
        - Very difficult to set up. There is no default database that can be used. The user has to create a database from scratch.
        - Not user friendly

- Open Context, Alexandria Archive Institute (opencontext.org)
    - Pros
        - Offers online access to raw data
        - Non-specialized fields that can be used for any data
        - Allows for files to be attached at several levels
        - Includes context
        - Free to use
    - Cons
        - Context does not offer a way to document how objects are related to one another
        - Cannot document precise provenience information
        - It is not a database management system. It does not offer a way to manage a collection
- tDAR, Digital Antiquity (www.tdar.org)
    - Pros
        - Offers online access to raw data
        - Allows associated files to be uploaded
        - Free to use
        - Compatible with Dublin Core
    - Cons
        - It is not a database management system. It does not offer a way to manage a collection

- Re:discovery, Rediscovery Software Inc. (www.rediscov.com)
    - Pros
        - Images are viewable on the form without clicking an external link
        - Incorporates a data dictionary which is editable
    - Cons
        - Who can add information or edit the data dictionary is not limited
        - Cannot document provenience or relate collection objects to other objects
- PastPerfect Museum Software 4, PastPerfect, Inc. (www.museumsoftware.com)
    - Pros
        - Can attach different file types
        - Different data entry forms available in the same database
        - Uses a data dictionary, which can be edited
    - Cons
        - Different data entry forms always available
        - Ability to attach images adds more money to the price of the software
        - Anyone can edit the data dictionary
        - Cannot associate collection objects with other objects

Appendix 4 – Application of the Data Model

Throughout this thesis, the data model has been the object of theoretical discussion. In order to test this model, I used a small sample of real data from three datasets from the same institution. The datasets are from the University of Kansas Archaeological Research Collection (Spreadsheet for Analysis NSF, Talking Crow Endscrapers, and Burntwood 2006) and represent raw data from the researcher collecting it. The data have not yet been entered into a database. The columns in the data model represent the fields that the researcher wanted to collect.

I used three rows from each spreadsheet (the actual data used is on the last page of this appendix). I attempted to take a representative sample using rows that seemed to have as many completed fields as possible. Each column was mapped (matched) to a field in the proposed data model. If a column was empty for all of the rows, I left the column out of the mapping. The mapped fields were used to identify the Specify tables that would be used. The relationship between the tables used by the data are shown below. I have put the first row from the three samples into a mockup of the data entry screen.

Spreadsheet for Analysis NSF

The first set of data are from "Spreadsheet for Analysis NSF." The data are separated into the following columns: SITE, ANALYSIS UNIT, CATALOG NUMBER, COMPLETENESS CATEGORY, RIM FORM, RIM HEIGHT-OUTSIDE, RIM HEIGHT-INSIDE, RIM ANGLE, LIP FORM, LIP THICKNESS, COLLAR THICKNESS, NECK THICKNESS, CHANNEL DEPTH, COLLAR PANEL SHAPE, COLLAR BASE SHAPE, DECORATION?, and HANDLE(S)?. The sample data are displayed in Spreadsheet 5 at the end of this section. In the Specify for Archaeology data model, the columns would be grouped into tables for COLLECTION OBJECT, LOCALITY, and ATTRIBUTE SET. The data would be placed in Specify as can be seen in Table 1.

Table 1 Mapping of columns from the "Spreadsheet for Analysis NSF" spreadsheet from the Archaeological Research Center in the Biodiversity Institute at the University of Kansas to the proposed data model.

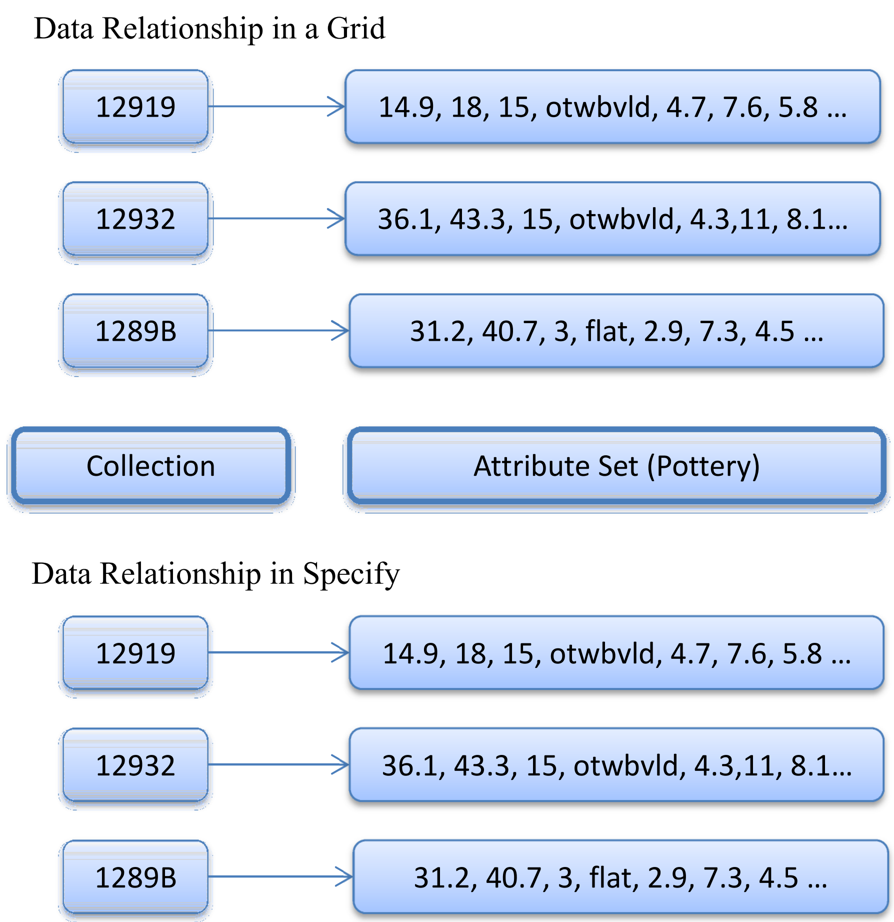
Spreadsheet for Analysis NSF		
Spreadsheet Column	**Specify Field**	**Specify Table**
Catalog number	Catalog Number	Collection Object
Site	Site Name	Locality
Analysis unit	Text 1	Attribute Set (Pottery)
Completeness category	Text 2	Attribute Set (Pottery)
Rim form	Text 3	Attribute Set (Pottery)
Rim height -outside	Text 4	Attribute Set (Pottery)
Rim height - inside	Text 5	Attribute Set (Pottery)
Rim angle	Text 6	Attribute Set (Pottery)
Lip form	Text 7	Attribute Set (Pottery)
Lip thickness	Text 8	Attribute Set (Pottery)
Collar thickness	Text 9	Attribute Set (Pottery)
Neck thickness	Text 10	Attribute Set (Pottery)
Collar height	Text 11	Attribute Set (Pottery)
Channel depth	Text 12	Attribute Set (Pottery)
Collar panel shape	Text 13	Attribute Set (Pottery)
Collar base shape	Text 14	Attribute Set (Pottery)
Decoration?	Yes/No 1	Attribute Set (Pottery)
Handle(s)?	Yes/No 2	Attribute Set (Pottery)

In this example, COLLECTION Object (CATALOG NUMBER is a field in the COLLECTION OBJECT) is associated with the ATTRIBUTE SET table. In this spreadsheet, there is only one ATTRIBUTE SET form required because there is only one collection object type. COLLECTION OBJECT and ATTRIBUTE SET have a one-to-one relationship so each COLLECTION OBJECT record will be associated with an ATTRIBUTE SET record.

There is a slight mismatch between the data model and the data. The data model requires LOCALITY information (SITE is a field in the LOCALITY table) be associated with COLLECTION

Object through the Collecting Event table. However, the “Spreadsheet for Analysis NSF” data do not contain any Collecting Event data. This is not a problem though because no data are required in the Collecting Event table. The table can be empty except for linking fields, such as a link to locality, which will not be seen by the user. So Locality can be linked to Collection Object through Collecting Event even if Collecting Event does not have data in it.

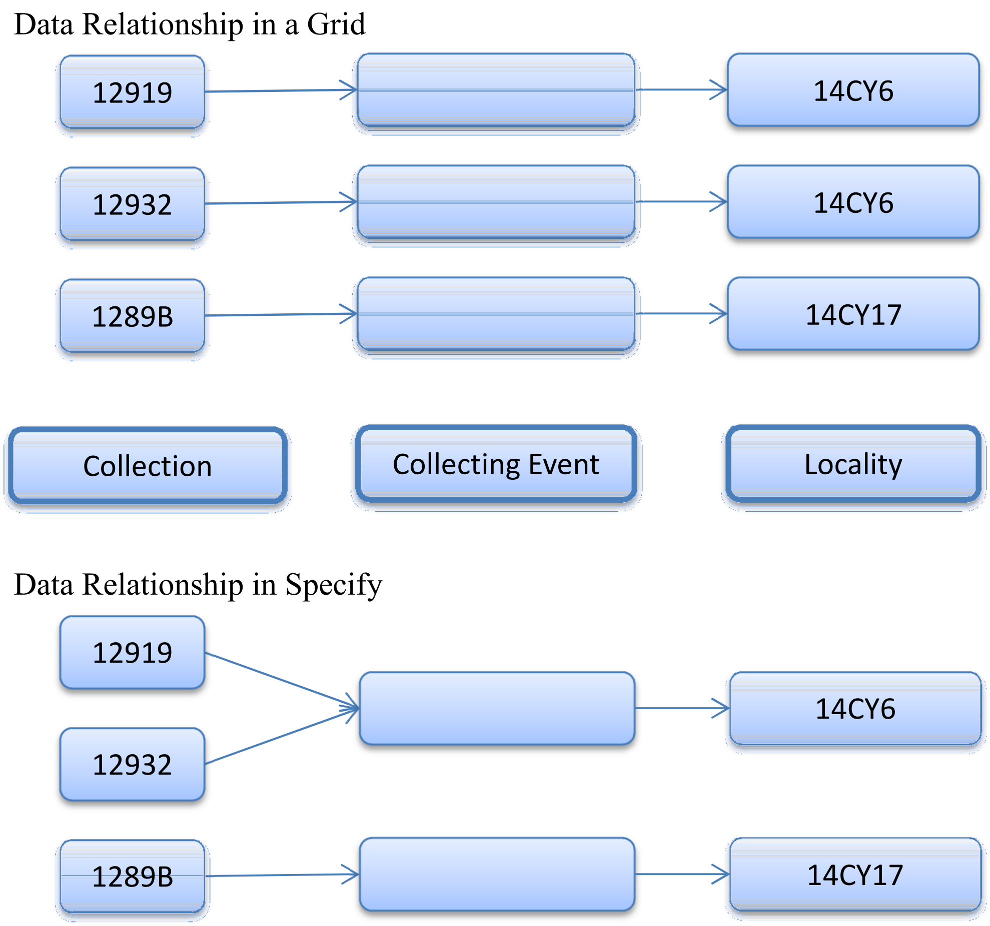
Figure 18 Spreadsheet for Analysis NSF Collection Object and Attribute Set data relationship.



In this example, there will be three Collection Object records created but only two Locality records created. Each Locality record will be related directly to a Collecting Event. Therefore two empty Collecting Event records will be created. One of the

COLLECTING EVENT records will be associated with one COLLECTION OBJECT record and another COLLECTING EVENT record will be associated with two COLLECTION OBJECT records.

Figure 19 Spreadsheet for Analysis NSF Collection Object, Collecting Event, and Locality data relationship.



Talking Crow Endscrapers

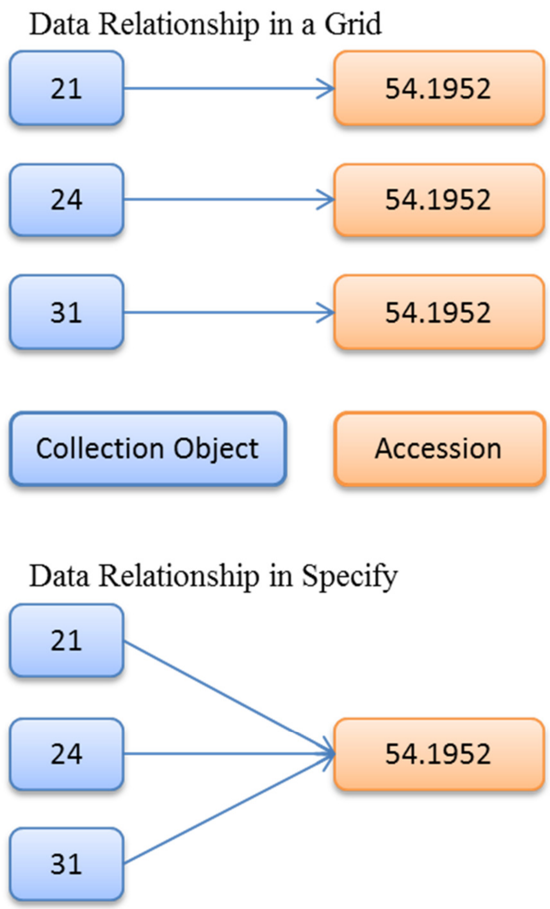
The first set of data are from "Talking Crow Endscrapers." The data are separated into the following columns: ACC#, SITE NAME, SITE NUMBER, CATALOG NUMBER, SPECIMEN NUMBER, YEAR, FEA#, OTHER PROVENIENCE, DEPTH/LEVEL, EX DATE, ARTIFACT TYPE, MAT TYPE, CON, BURN, L, W, T, WT, ULTRA-VIOLET SHORT WAVE LENGTH, and ULTRA-VIOLET LONG WAVE-LENGTH. The sample data are displayed in the second spreadsheet at the end of this section. In the Specify for Archaeology data model, the columns would be grouped into tables for ACCESSION, COLLECTION OBJECT, COLLECTING EVENT, LOCALITY, PROVENIENCE, FEATURE, and ATTRIBUTE SET. The data would be placed in Specify as can be seen in Table 2.

Table 2 Mapping of columns from the "Talking Crow Endscrapers" spreadsheet from the Archaeological Research Center in the Biodiversity Institute at the University of Kansas to the proposed data model.

Talking Crow Endscrapers		
Spreadsheet Column	**Specify Field**	**Specify Table**
Acc#	Accession Number	Accession
Site Name	Site Name	Locality
Site Number	Site Code	Locality
Fea#	Feature Number	Feature
Catalog Number	Catalog Number	Collection Object
Specimen Number	Alt Catalog Number	Collection Object
Year	Cataloged Date	Collection Object
Catalog Remarks	Remarks	Collection Object
Ex Date	Excavation Date	Collecting Event
Other Provenience	Trench	Provenience
Depth/ Level	Level	Provenience
Artifact Type	Object Type	Attribute Set (Stone)
Mat Type	Material	Attribute Set (Stone)
Con	Text 1	Attribute Set (Stone)
Burn	Yes/No 1	Attribute Set (Stone)
L	Number 1	Attribute Set (Stone)
W	Number 2	Attribute Set (Stone)
T	Number 3	Attribute Set (Stone)
Wt	Number 4	Attribute Set (Stone)
Ultra-violet Short wave-length	Text 1	Treatment Event
Ultra-violet long wave-length	Text 2	Treatment Event

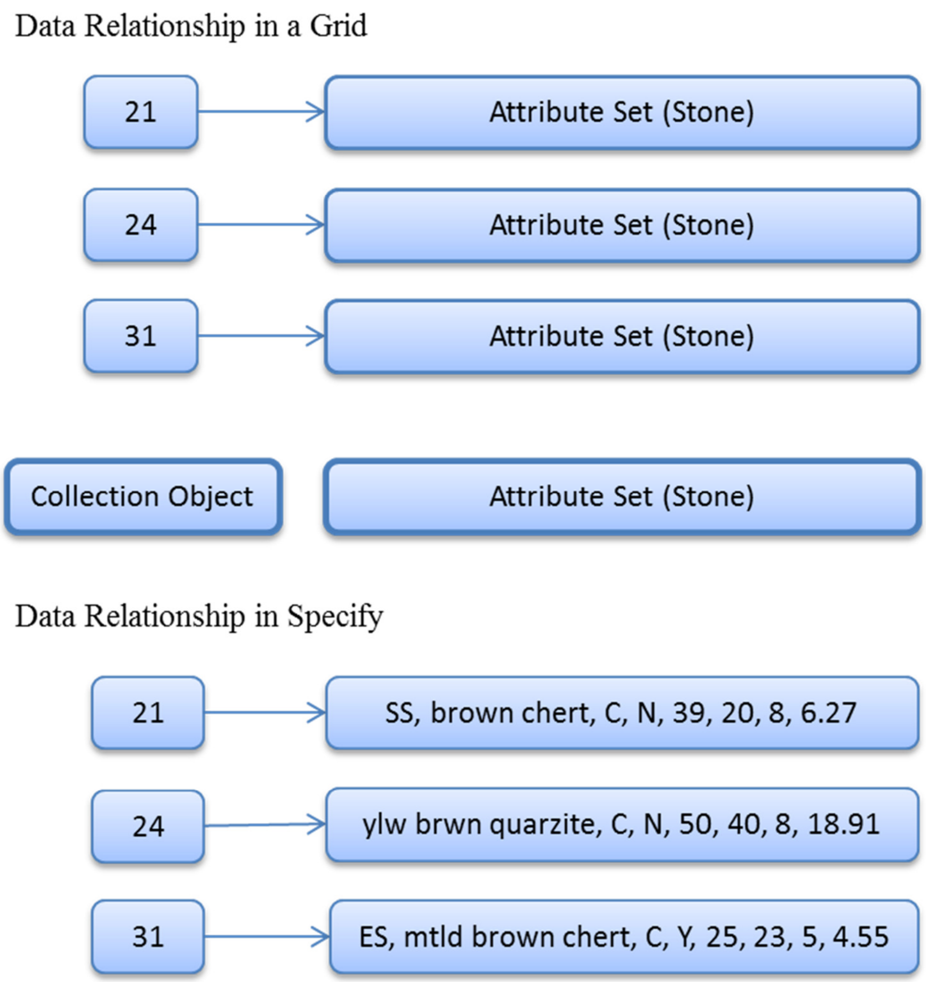
Three COLLECTION OBJECT records would be created. The records would consist of the CATALOG NUMBER, SPECIMEN NUMBER, YEAR, and CATALOG REMARKS. All three COLLECTION OBJECT records were entered into the Archaeological Research Center in the same accession, 54.1952. All three of the COLLECTION OBJECT records will associate with a single ACCESSION record.

Figure 20 Talking Crow Endscrapers Collection Object and Accession data relationship.



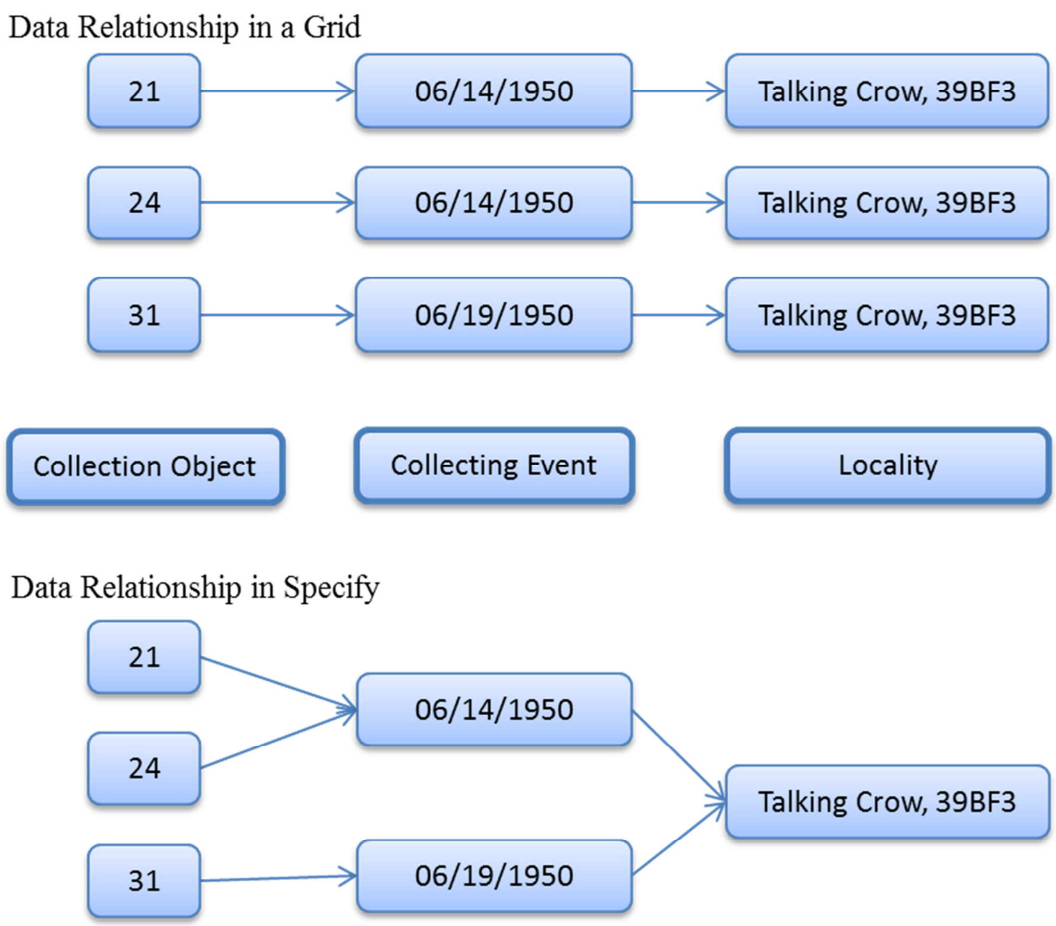
There is only one Attribute Set form required because there is only one collection object type. The Attribute Set form would consist of fields for Artifact Type, Mat Type, Con, Burn, L, W, T, and Wt. Though only one form is required, each Collection Object record will be associated with a different Attribute Set record. There will be three Attribute Set records created in this example.

Figure 21 Talking Crow Endscrapers Collection Object and Attribute Set data relationship.



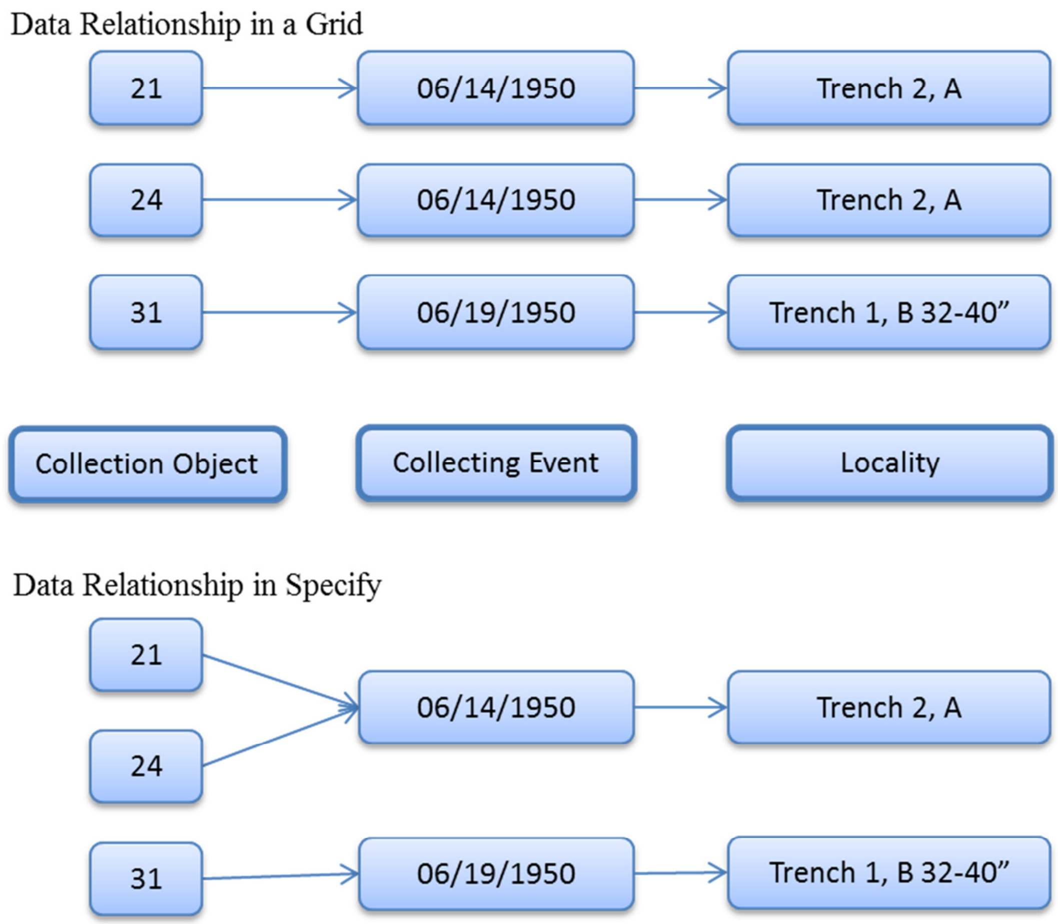
In this example, Collecting Event only consists of the excavation date (Ex Date). The date is the same for two records and different for a third. Therefore two records are created, one with the first date and one with the second. Two Collection Object records would be related to one of the Collecting Event records and the third Collection Object record would be related to a separate Collecting Event record. Both of the Collecting Event records relate to a single Locality record, which in this example consists of the Site Name and Site Number. Since the Site Name and Site Number is the same for all three records, only one Locality is required.

Figure 22 Talking Crow Endscrapers Collection Object, Collecting Event, and Locality data relationship.



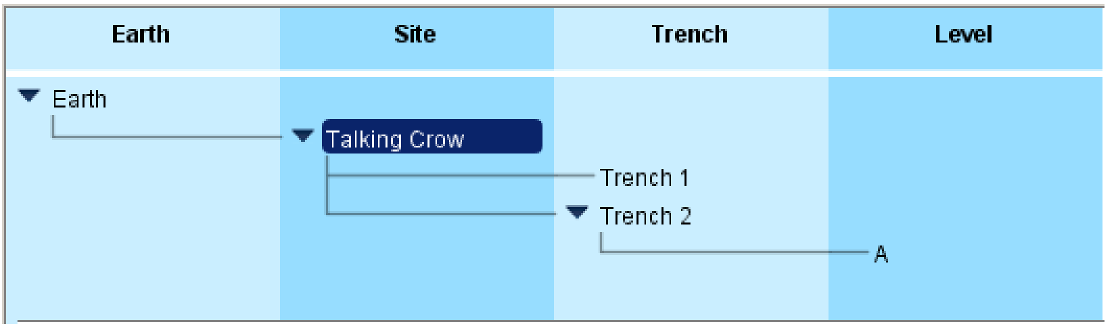
Further locality information is provided through the Other Provenience, Depth/Level, and Fea# fields. Other Provenience and Depth/Level would be documented in the Stratigraphy table. Stratigraphy is associated to the Collection Object record through the Collecting Event table. Two of the records have the same Other Provenience and Depth/Level and the third record has different data for those fields. The two records that share the Stratigraphy data also share the same Collecting Event data. The two Collecting Event records will therefore be directly related to a Stratigraphy record.

Figure 23 Talking Crow Endscrapers Collection Object, Collecting Event, and Stratigraphy data relationship.



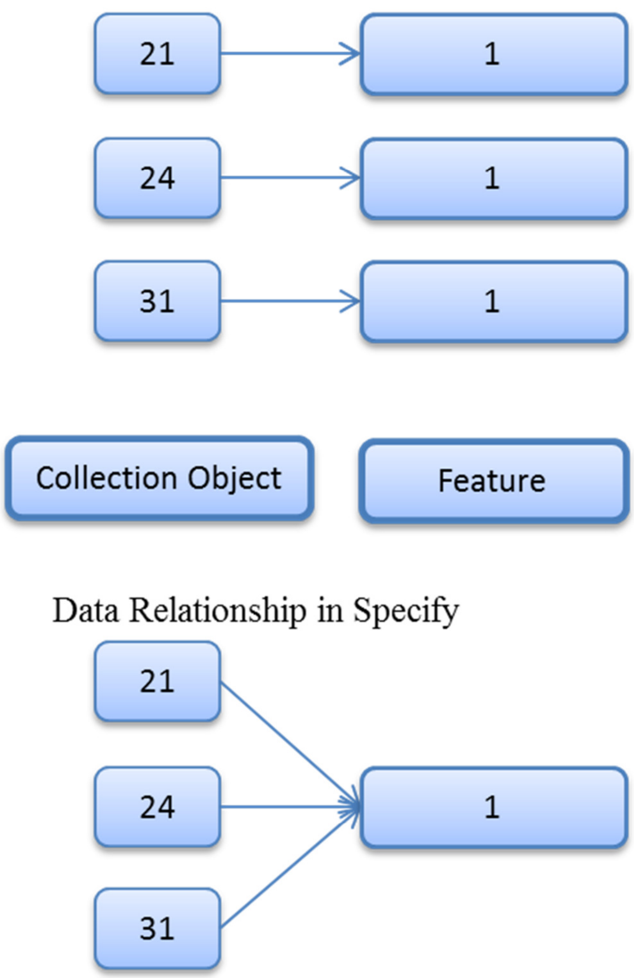
Because STRATIGRAPHY data are documented as a hierarchy, information about each level of the stratigraphy can be documented. Information such as level elevation and matrix can be documented for each provenience level. In the example above, three records would be created; one each for Site, Trench, and Level. The user could decide to use more or fewer levels as they wanted. Differing information could be documented for Trench 2 and for Level A if the user had more information.

Figure 24 Sample screen shot of a Stratigraphy hierarchy.



FEA# will be documented in the FEATURE Table. Since the documented FEATURE information is limited to a single shared FEATURE NUMBER, only one FEATURE record will be created.

Figure 25 Talking Crow Endscrapers Collection Object and Feature data relationship.



Burntwood 2006

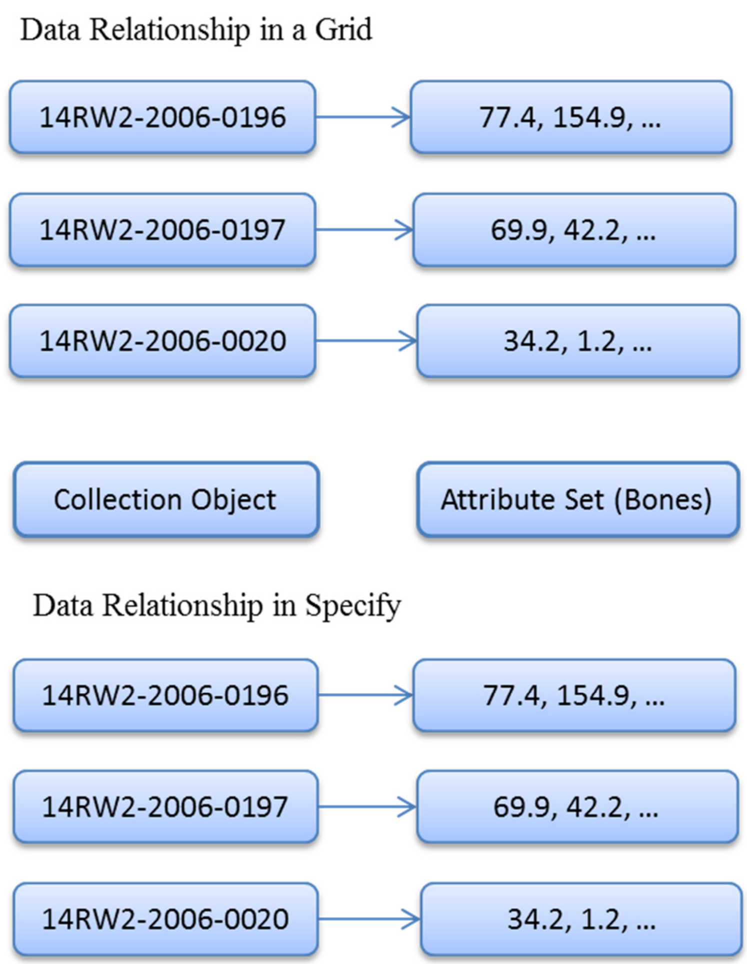
The third set of data are from "Burntwood 2006". The data are separated into the following columns: CAT #, OLD #, NOTES, SPECIES, ELEMENT, CODE, SIDE, PORTION, COMPLETE %, MX L MM, WT GM, BURNED, MODIFIED?, JUV?, COLLECTION METHOD, EXCAVATOR, and PROVENANCE. The sample data are displayed in Spreadsheet 3 at the end of this section. In the Specify for Archaeology data model, the columns would be grouped into tables for ACCESSION, COLLECTION OBJECT, COLLECTING EVENT, LOCALITY, PROVENIENCE, FEATURE, and ATTRIBUTE SET. The data would be placed in Specify as can be seen in Table 3.

Table 3 Mapping of columns from the "Burntwood 2006" spreadsheet from the Archaeological Research Center in the Biodiversity Institute at the University of Kansas to the proposed data model.

Burntwood 2006		
Spreadsheet Column	**Specify Field**	**Specify Table**
Cat #	Catalog Number	Collection Object
old #	Alt Catalog Number	Collection Object
Notes	Remarks	Collection Object
Species	Name	Classification
Element	Description	Preparation
Code	Text 1	Preparation
Side	Text 2	Preparation
Portion	Text 3	Preparation
Complete %	Text 4	Preparation
Mx L mm	Number 1	Attribute Set (Bones)
Wt gm	Number 2	Attribute Set (Bones)
Burned	Yes/No 1	Attribute Set (Bones)
Modified?	Yes/No 2	Attribute Set (Bones)
Juv?	Yes/No 3	Attribute Set (Bones)
Collection method	Collection Method	Collecting Event
Excavator	Excavator	Collecting Event
Provenance	Verbatim Provenience	Collecting Event

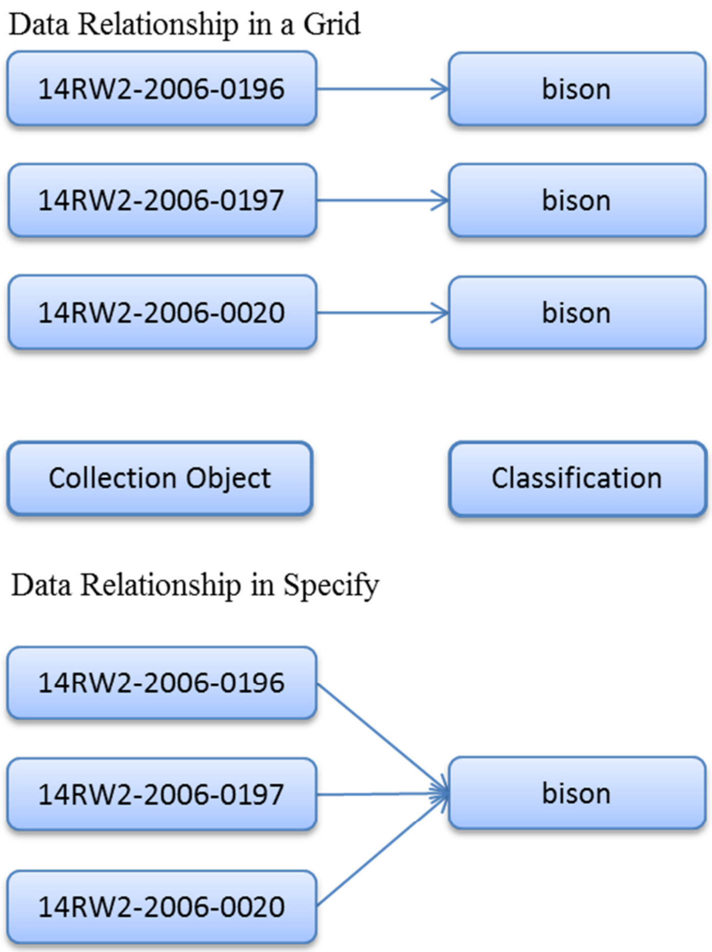
Three COLLECTION OBJECT records would be created. The records would consist of the CAT #, OLD #, and NOTES. Only one ATTRIBUTE SET form required because there is only one collection object type. The ATTRIBUTE SET form would consist of fields for MX L MM, WT GM, BURNED, MODIFIED?, and JUV?. As with the previous spreadsheet, although only one form is required, each COLLECTION OBJECT record will be associated with a different ATTRIBUTE SET record. There will be three ATTRIBUTE SET records created in this example.

Figure 26 Burntwood 2006 Collection Object and Attribute Set data relationship.



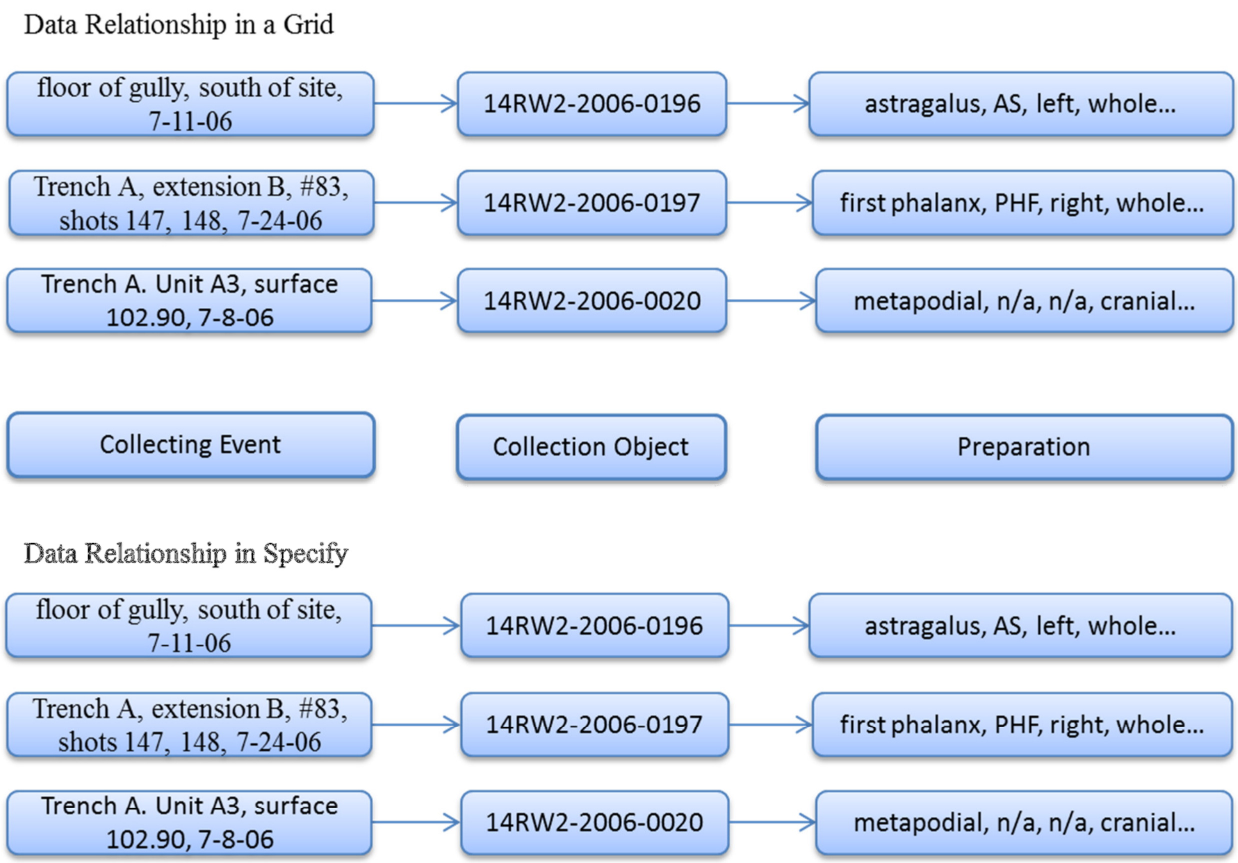
Further description is provided through the CLASSIFICATION table. The CLASSIFICATION record in this example consists of the SPECIES field. All three COLLECTION OBJECT records will be related to the single CLASSIFICATION record.

Figure 27 Talking Crow Endscrapers Collection Object, Classification, and Typology data relationship.



A description of the collection of the object is provided through the Collecting Event table. Collecting Event consists of the Collection method, Excavator, and Provenance. Although the Collecting Method and Excavator data are the same for all three collecting event records, the Provenience is different for all three. Because the Provenience is unique in each of the records, three Collecting Event records will be created. Each collection object record will be directly related to a unique collecting event record.

Figure 28 Talking Crow Endscrapers Collecting Event, Collection Object, and Preparation data relationship.



Three PREPARATION records will be created. The PREPARATION form will consist of ELEMENT, CODE, SIDE, PORTION, AND COMPLETE %. The information is different for all three records so three separate records will be created. Each COLLECTION OBJECT will be directly related to a single PREPARATION record.

Combined Spreadsheets

So far we have looked at each spreadsheet separately. However, the point of the data model is to accommodate all of the data that would be entered into the database. Therefore, to properly examine the data model, we must look at how the data will be combined in the data model. The first row of data from all three spreadsheets was entered into a mockup of the data entry form for the data model. These three rows created three pretend records in the database.

If all three spreadsheets are imported as is into the same database, all of the fields and tables mentioned in the mappings for those spreadsheets would have to be available to all users. As was described in this thesis, all of the tables will be shared except for a few artifact type specific tables such as the Attribute Set table, which is the only table in this example that is artifact type specific.

Table 4 Mapping of the fields from all three of the spreadsheets to the proposed data model.

Spreadsheet Column	Specify Field	Specify Table
Acc#	Accession Number	Accession
Mx L mm	Number 1	Attribute Set (Bones)
Wt gm	Number 2	Attribute Set (Bones)
Burned	Yes/No 1	Attribute Set (Bones)
Modified?	Yes/No 2	Attribute Set (Bones)
Juv?	Yes/No 3	Attribute Set (Bones)
Analysis unit	Text 1	Attribute Set (Pottery)
Completeness category	Text 2	Attribute Set (Pottery)
Rim form	Text 3	Attribute Set (Pottery)
Rim height -outside	Text 4	Attribute Set (Pottery)
Rim height - inside	Text 5	Attribute Set (Pottery)
Rim angle	Text 6	Attribute Set (Pottery)
Lip form	Text 7	Attribute Set (Pottery)
Lip thickness	Text 8	Attribute Set (Pottery)
Collar thickness	Text 9	Attribute Set (Pottery)
Neck thickness	Text 10	Attribute Set (Pottery)
Collar height	Text 11	Attribute Set (Pottery)
Channel depth	Text 12	Attribute Set (Pottery)
Collar panel shape	Text 13	Attribute Set (Pottery)
Collar base shape	Text 14	Attribute Set (Pottery)
Decoration?	Yes/No 1	Attribute Set (Pottery)
Handle(s)?	Yes/No 2	Attribute Set (Pottery)
Artifact Type	Object Type	Attribute Set (Stone)
Mat Type	Material	Attribute Set (Stone)
Con	Text 1	Attribute Set (Stone)
Burn	Yes/No 1	Attribute Set (Stone)

L	Number 1	Attribute Set (Stone)
W	Number 2	Attribute Set (Stone)
T	Number 3	Attribute Set (Stone)
Wt	Number 4	Attribute Set (Stone)
Ex Date	Excavation Date	Collecting Event
Collection method	Collection Method	Collecting Event
Excavator	Excavator	Collecting Event
Provenance	Verbatim Provenience	Collecting Event
Catalog Number	Catalog Number	Collection Object
Alt Catalog Number	Alt Catalog Number	Collection Object
Year	Cataloged Date	Collection Object
Remarks	Remarks	Collection Object
Fea#	Feature Number	Feature
Site Name	Site Name	Locality
Site Number	Site Code	Locality
Element	Description	Preparation
Code	Text 1	Preparation
Side	Text 2	Preparation
Portion	Text 3	Preparation
Complete %	Text 4	Preparation
Other Provenience	Trench	Provenience
Depth/ Level	Level	Provenience
Ultra-violet Short wave-length	Text 1	Treatment Event
Ultra-violet long wave-length	Text 2	Treatment Event
Species	Name	Typology

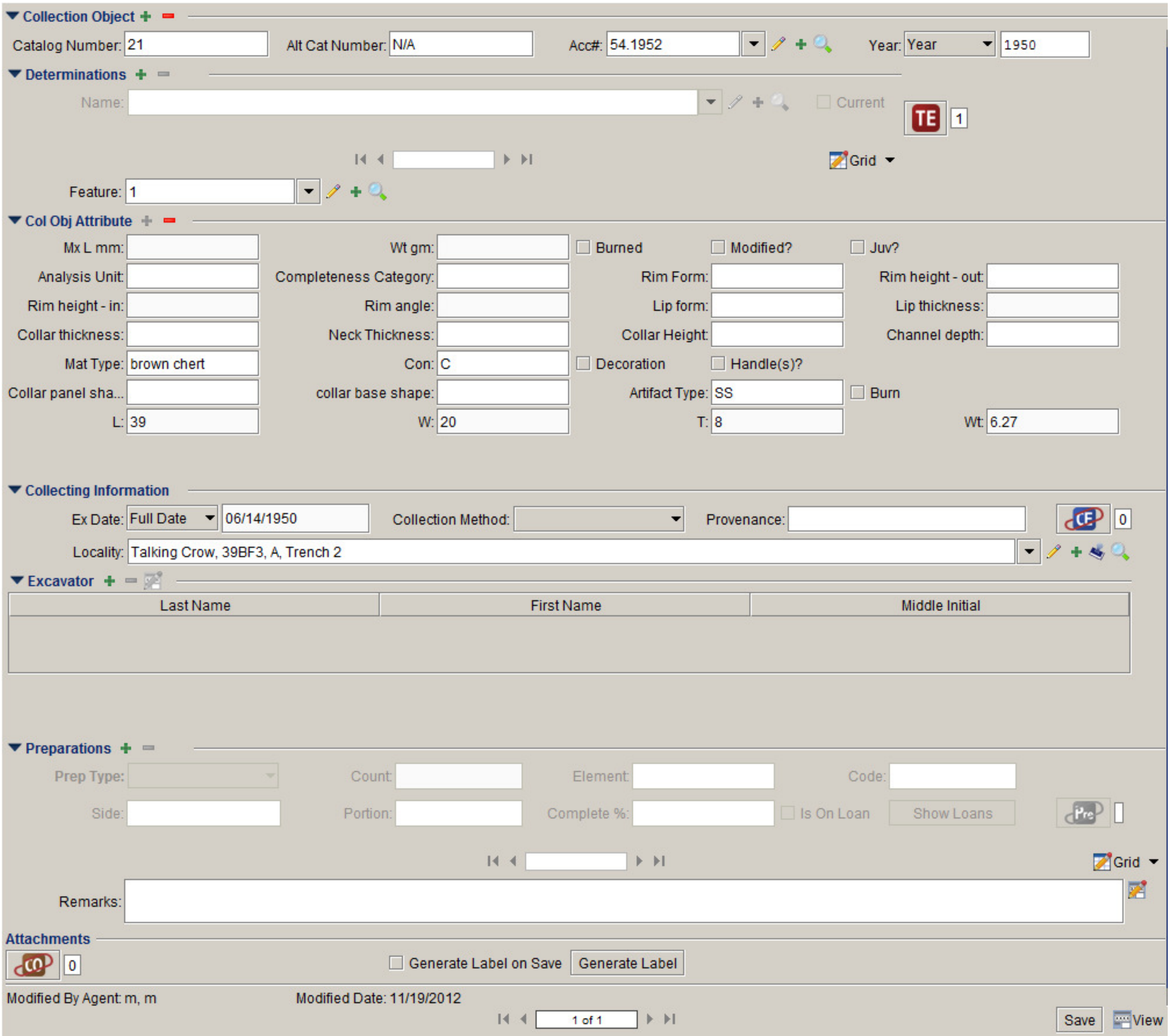
The field list gets long because most fields are not duplicated. There are four fields that are duplicated on multiple spreadsheets. The only field that is duplicated on all three spreadsheets is CATALOG NUMBER (also called CAT #). The other three fields are duplicated on two of the three spreadsheets. Those fields are: ALT CATALOG NUMBER (called SPECIMEN NUMBER and OLD #), COLLECTION OBJECT REMARKS (called CATALOG REMARKS and NOTES), and SITE NAME (called SITE NAME and SITE).

For the purposes of this discussion, I will use the Specify field name rather than the caption the user used for the four fields that are on multiple spreadsheets in order to reduce confusion. If this were a real database, the user would have to decide on a single field caption to use for each field. The decided caption would be the field caption regardless of the spreadsheet column name.

In other database systems such as Past Perfect and Re:Discovery, if the database manager would have two options to handle this data. They would either have to eliminate fields in order to simplify the data entry forms or they would have to expand the data entry form to allow for all of the fields used by data contributors. If my data model were to handle the data in the way these systems would handle the data without excluding columns, the data entry form would look like Image 28 below.[4]

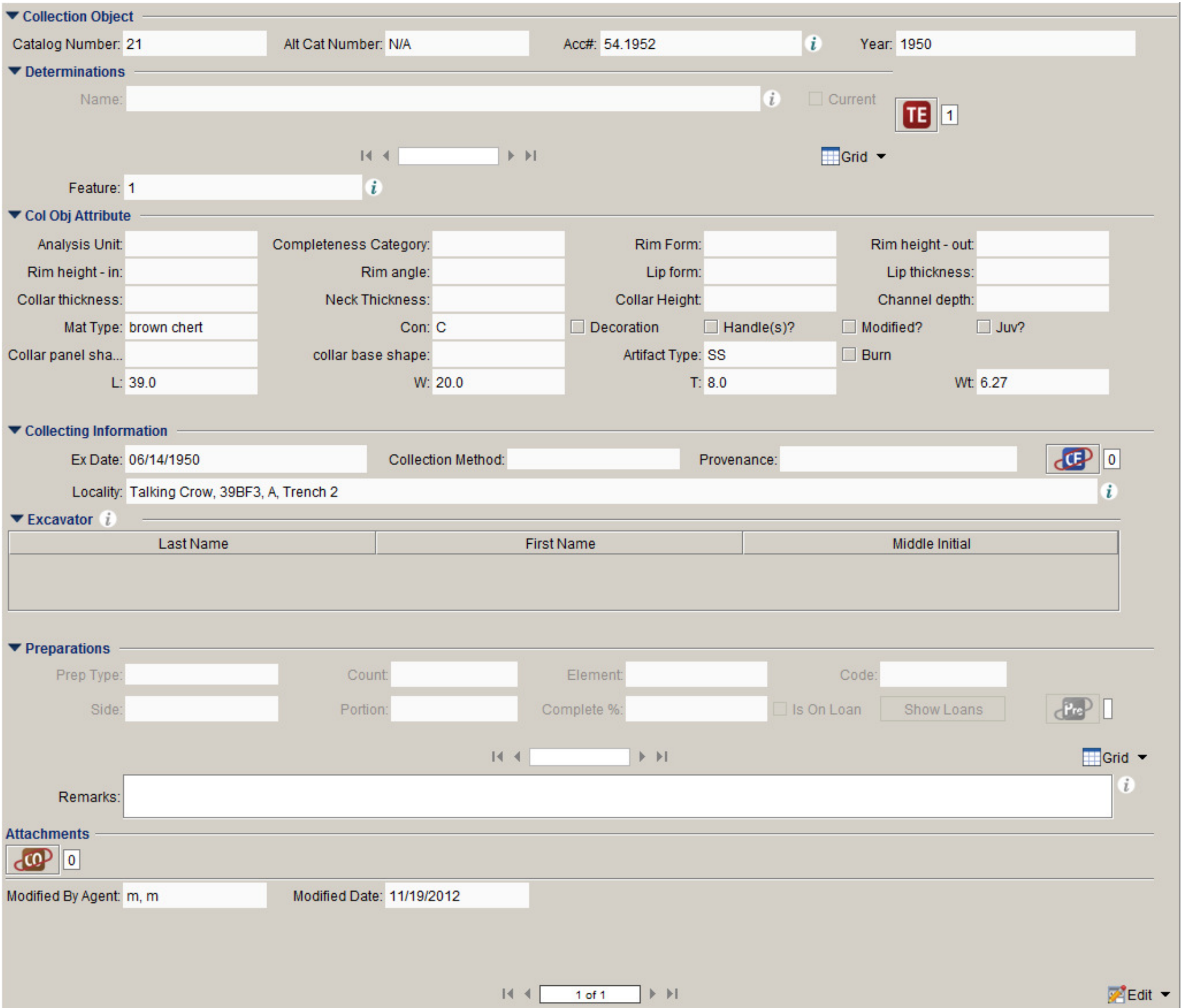
[4]I used Specify to demonstrate how the other programs would handle the data because almost all of the programs described in Chapter 3 except tDAR and Open Context are (Microsoft Access 2003, Re:Discovery, PastPerfect, and KeEmu) are either expensive to purchase and /or require training in order to use.

Figure 29 Sample screen shot of the Collection Object data entry form using the data relationship in Past Perfect or Re:Discovery with the first record from the Talking Crow Endscrapers spreadsheet.



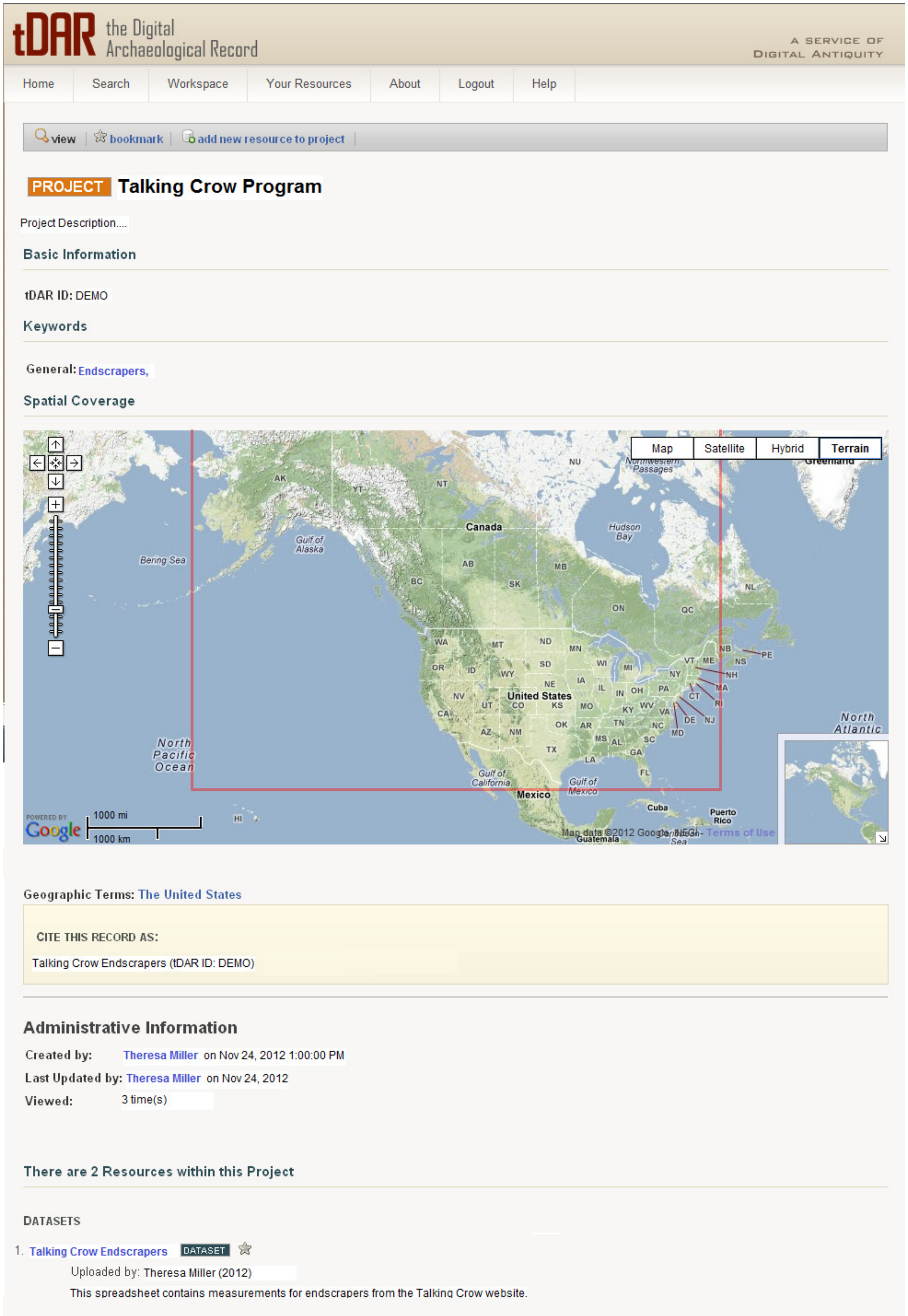
In image 27, I made each of the fields a unique field in the form, without evaluating to see if the fields appear duplicated. If I edit the table to remove fields that appear to be duplicates, the form appears as image 28. Image 28 is a bit more streamlined than Image 27 but there are still 12 fields the user has to bypass in order to get to the first relevant fields for this record from the Talking Crow Endscrapers spreadsheet.

Figure 30 Sample screen shot of the edited Collection Object data entry form using the data relationship in Past Perfect or Re:Discovery with the first record from the Talking Crow Endscrapers spreadsheet.



Alternatively, tDAR would not need to adapt its data entry form because the user does not enter data into a data entry form. tDAR allows the user to describe the data then attach the actual data as is demonstrated in image 30 below. As was discussed in Chapter 3, programs like tDAR and Open Context are not Database Management Systems like Specify so there are different requirements for its data entry, searching, and display capabilities.

Figure 31 Sample screen shot of the Talking Crow Endscrapers dataset in the tDAR.



With my data model, the database manager would have multiple options, she could set standards for data entry and only allow certain fields in order to streamline the database, she could create a single ATTRIBUTE SET to account for all of the attributes (this would appear like the form in the other database example), she could create an ATTRIBUTE SET form for each of the different spreadsheets, or she could create a few ATTRIBUTE SET forms that will account for the different artifact types.

In the combined database, the basic form would not change regardless of which spreadsheet's data were being entered. Where the form would differ is in the ATTRIBUTE SET. As can be seen in Images 30, 31, and 32, the basic collection object form layout is the same for all of the records. The same fields are offered, in the same order, with the same captions. Although you can only see the LOCALITY form in Image 32, it is the same layout for all three records as well.

In this example I decided to customize the ATTRIBUTE SET form for all three data types. I could have chosen to combine the attributes of two or all three of the records into a single form, but I wanted to show the versatility. If the data model were put into production, the user would have to make the same decision on how many ATTRIBUTE SET forms to create for his or her own database.

Figure 32 Sample screen shot of the Collection Object and Attribute Set forms with the first row of data from the Burntwood 2006 spreadsheet.

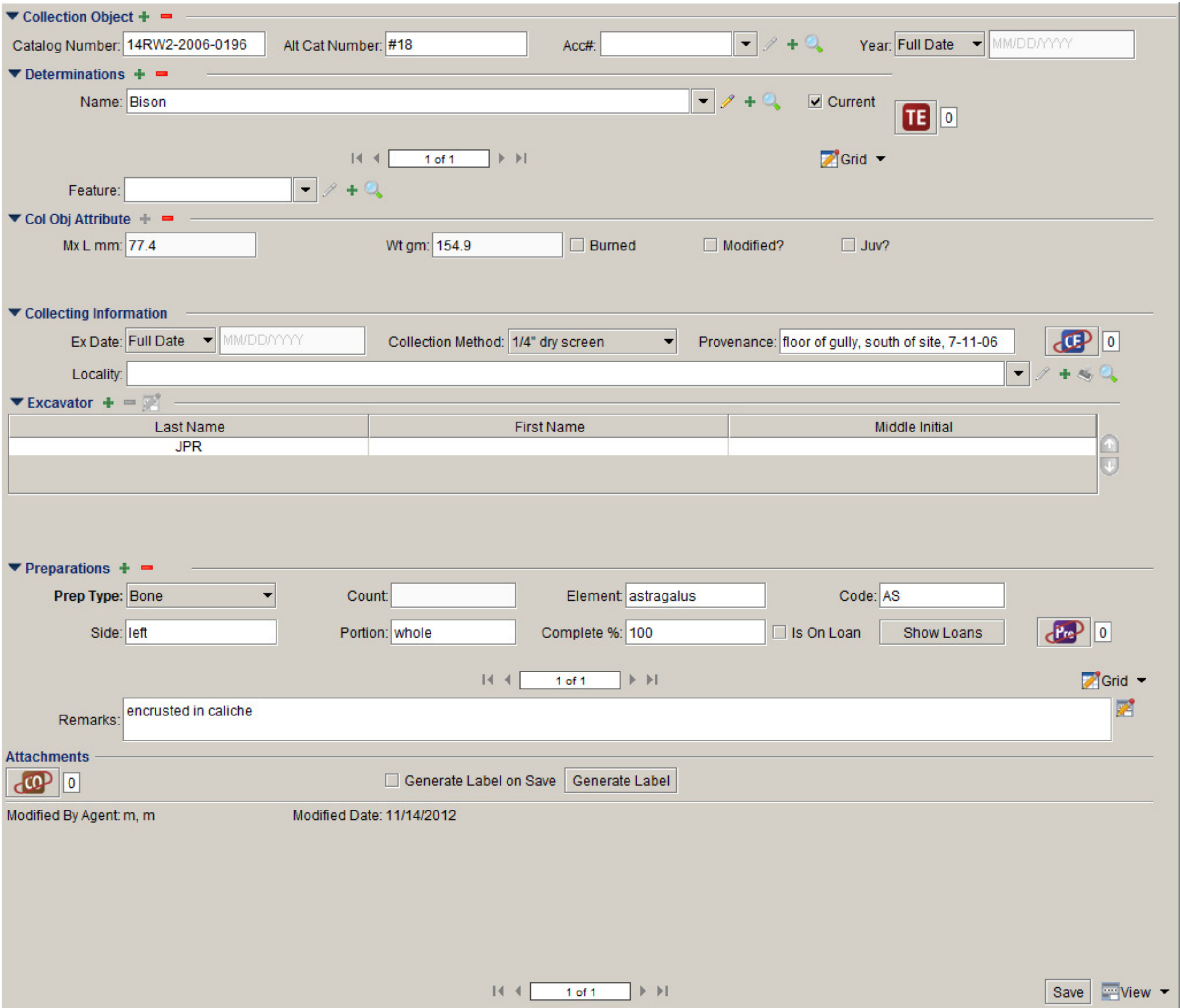
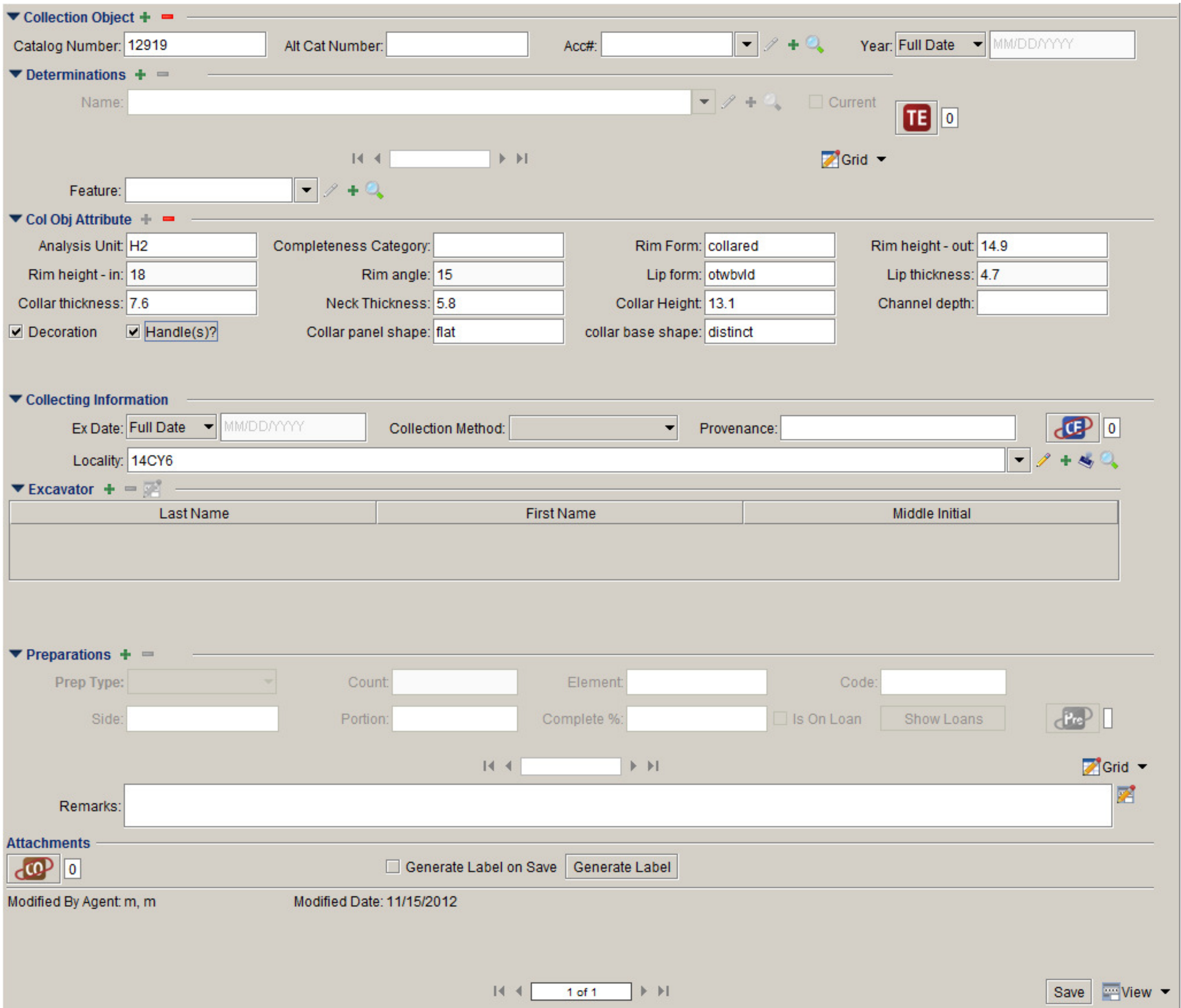


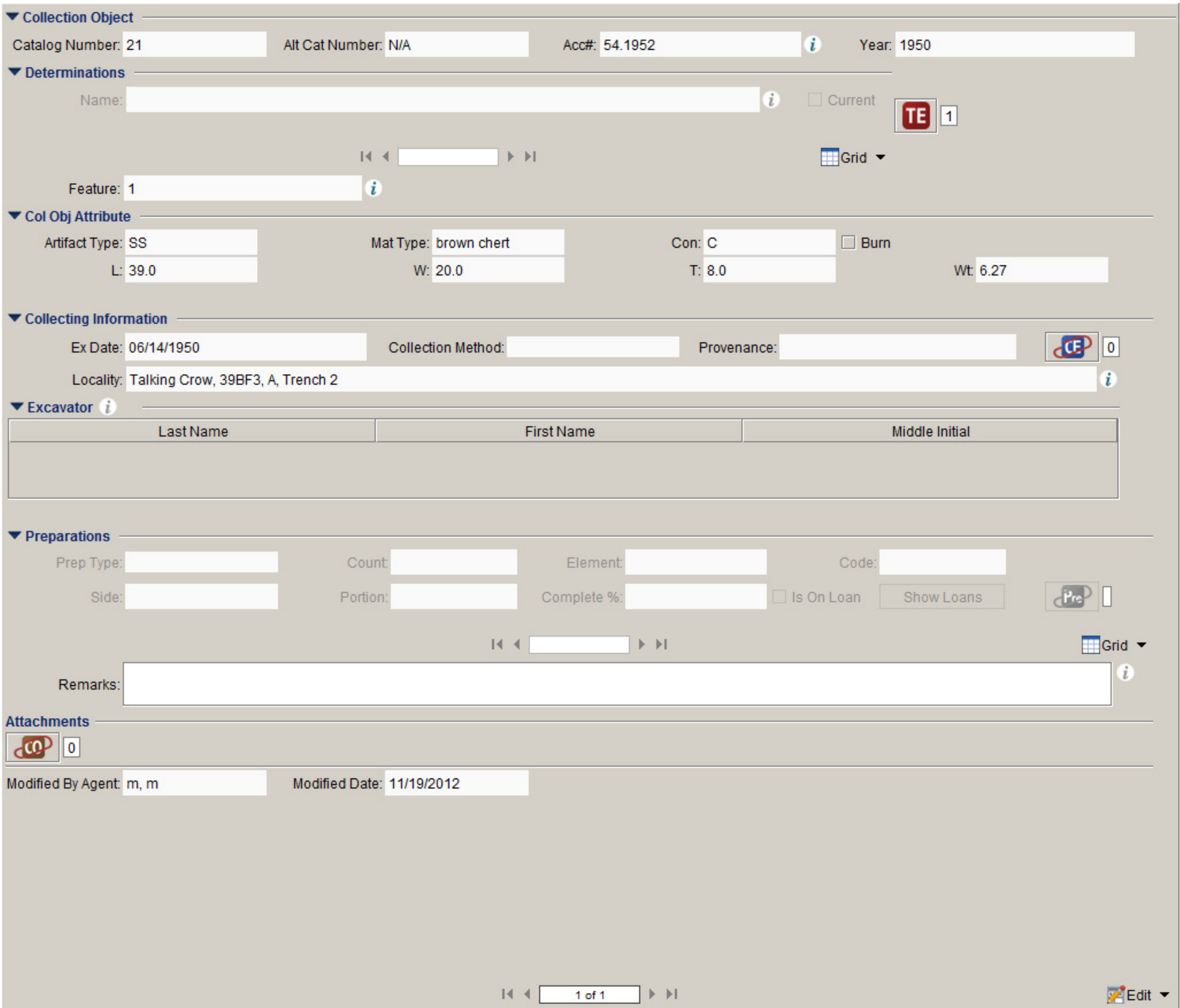
Figure 33 Sample screen shot of the Collection Object and Attribute Set forms with the first row of data from the Spreadsheet for Analysis NSF spreadsheet.

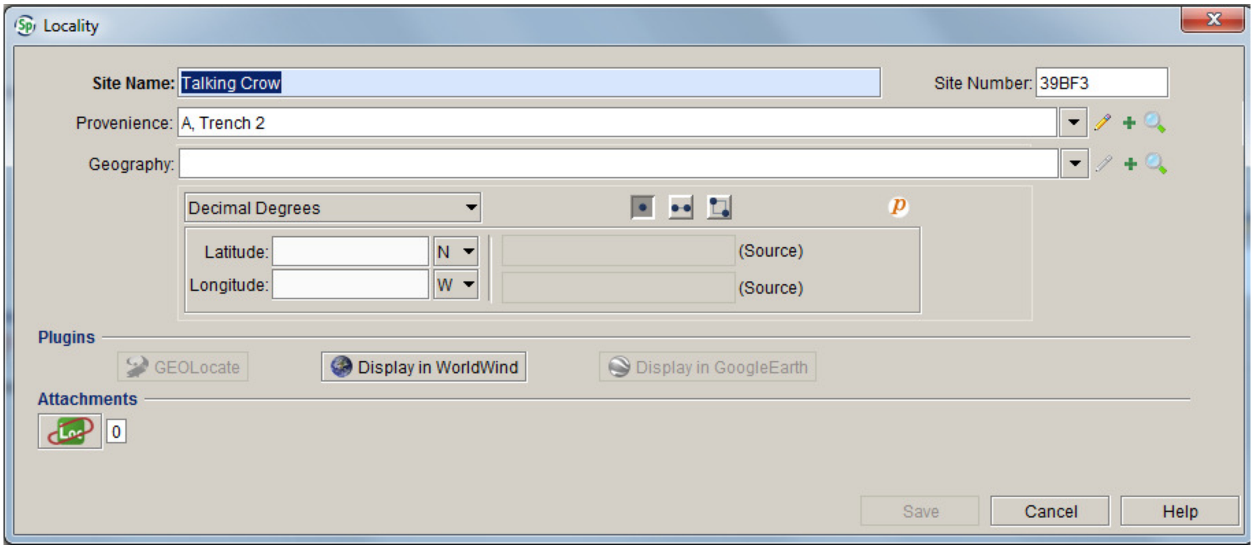


The Talking Crow Endscrapers spreadsheet presented in Figure32 demonstrates the second advantage of the data model, how it handles LOCALITY data. The spreadsheet demonstrates handling LOCALITY data through the FEATURE table and the PROVENIENCE table. Although the spreadsheet uses the tables it does not fully demonstrate the versatility of the tables.

The example above places the object in Feature #1. The spreadsheet only documents FEATURE NUMBER but more information about the feature could be documented such as TYPE, MEASUREMENTS, PROVENIENCE, etc. If the feature were within another feature, that could be documented in this data model as well.

Figure 34 Sample screen shot of the Collection Object, Attribute Set, Treatment Event, and Locality forms with the first row of data from the Talking Crow Endscrapers spreadsheet.





The data that was used is analytical data straight from the field, collected by different people. This is the strength and the weakness of the data. The strength is that it shows the variability of the data model in a real world experience. Collections may include data from many sources. Those sources may choose to document different data. The example showed how that variation in data could be accounted for.

The weakness of the data are that the data have not been reviewed by a collection manager. A collection manager may choose to exclude some of the data in order to standardize the data that is entered into the database. Further, because it is analytic data it does not include administrative data such as DATE CATALOGED, CATALOGER, ACCESSION NUMBER, etc. Also, the data did not contain information on context. Context has been a significant part of my discussion and my rationale why a new DBMS is needed, because the commercially available systems do not offer a way to associate collection objects.

Regardless of the limitations, the example shows the most significant aspects of the data model; the variability required to document the distinct data needs of an archaeology database and the ability to document locality information such as feature and provenience. As was discussed previously in this thesis, the administrative requirements do not vary much between biological and archaeological collections. Thus, the data model will document the administrative data for an archaeological collection in the same effective manner as it does for biological collections.

Table 5 Raw Data from the Spredsheet for Analysis NSF spreadsheet, Talking Crow Endscrapers, and Burntwood 2006 spreadsheets.

Spreadsheet for Analysis NSF

Site	Analysis unit	Catalog number	Completeness category	Rim form	Rim height -outside	Rim height - inside	Rim angle	Lip form
14CY6	H2	12919		collared	14.9	18	15	otwbvld
14CY6	H2	12932		collared	36.1	43.3	15	otwbvld
14CY17	70	12589B	LRB	collared	31.2	40.7	3	flat
Lip thickness	**Collar thickness**	**Neck thickness**	**Collar height**	**Channel depth**	**Collar panel shape**	**Collar base shape**	**Decoration?**	**Handle(s)?**
4.7	7.6	5.8	13.1		flat	distinct	TRUE	TRUE
4.3	11	8.1	21.5		concave	distinct	TRUE	FALSE
2.9	7.3	4.5	21.9		flat	distinct	FALSE	FALSE

Talking Crow Endscrapers

Acc#	Site Name	Site Number	Catalog Number	Specimen Number	Year	Fea#	Other Provenience	Depth/ Level	Catalog Remarks	
54.1952	Talking Crow	39BF3	21	N/A	1950	1	Trench 2	A		
54.1952	Talking Crow	39BF3	24	N/A	1950	1	Trench 2	A		
54.1952	Talking Crow	39BF3	31	N/A	1950	1	Trench 1	B 32-40”		
Ex Date	**Artifact Type**	**Mat Type**	**Con**	**Burn**	**L**	**W**	**T**	**Wt**	**Ultra-violet Short wave-length**	**Ultra-violet long wave-length**
06/14/50	SS	brown chert	C	N	39	20	8	6.27	none	none
06/14/50		ylw brwn quartzite	C	N	50	40	8	18.91	none	red
06/19/50	ES	mtld brown chert	C	Y	25	23	5	4.55	none	white incl

Burntwood 2006

Cat #	old #	Species	Element	Code	Side	Mx L mm	Wt gm	Portion	Juv?	Complete %
14RW2-2006-0196	#18	bison	astragalus	AS	left	77.4	154.9	whole	no	100
14RW2-2006-0197	#35	bison	first phalanx	PHF	right	69.9	42.2	whole	no	100
14RW2-2006-0020	n/a	bison	metapodial	n/a	n/a	34.2	1.2	cranial shaft fragment	n/a	n/a
Collection method	**Burned**	**Modified?**	**Excavator**	**Provenance**				**Notes**		
1/4" dry screen	n/a	n/a	JPR	floor of gully, south of site, 7-11-06				encrusted in caliche		
1/4" dry screen	n/a	n/a	JPR	Trench A, extension B, #83, shots 147, 148, 7-24-06						
1/4" dry screen	no	no	JPR	Trench A. Unit A3, surface 102.90, 7-8-06						

Appendix 5 – The Data Model

Figure 35 Data Model

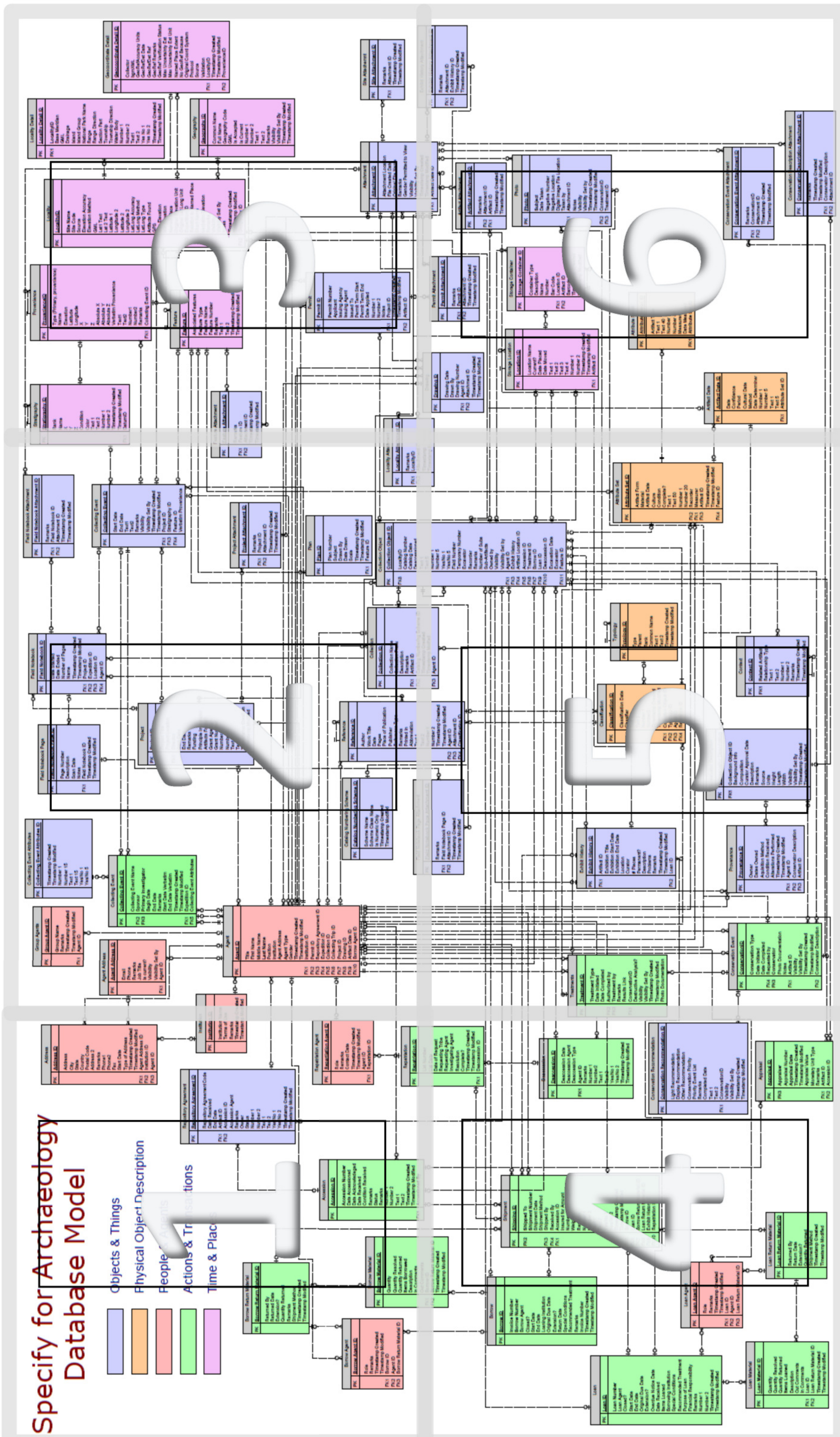


Figure 36 Section 1 of Data Model

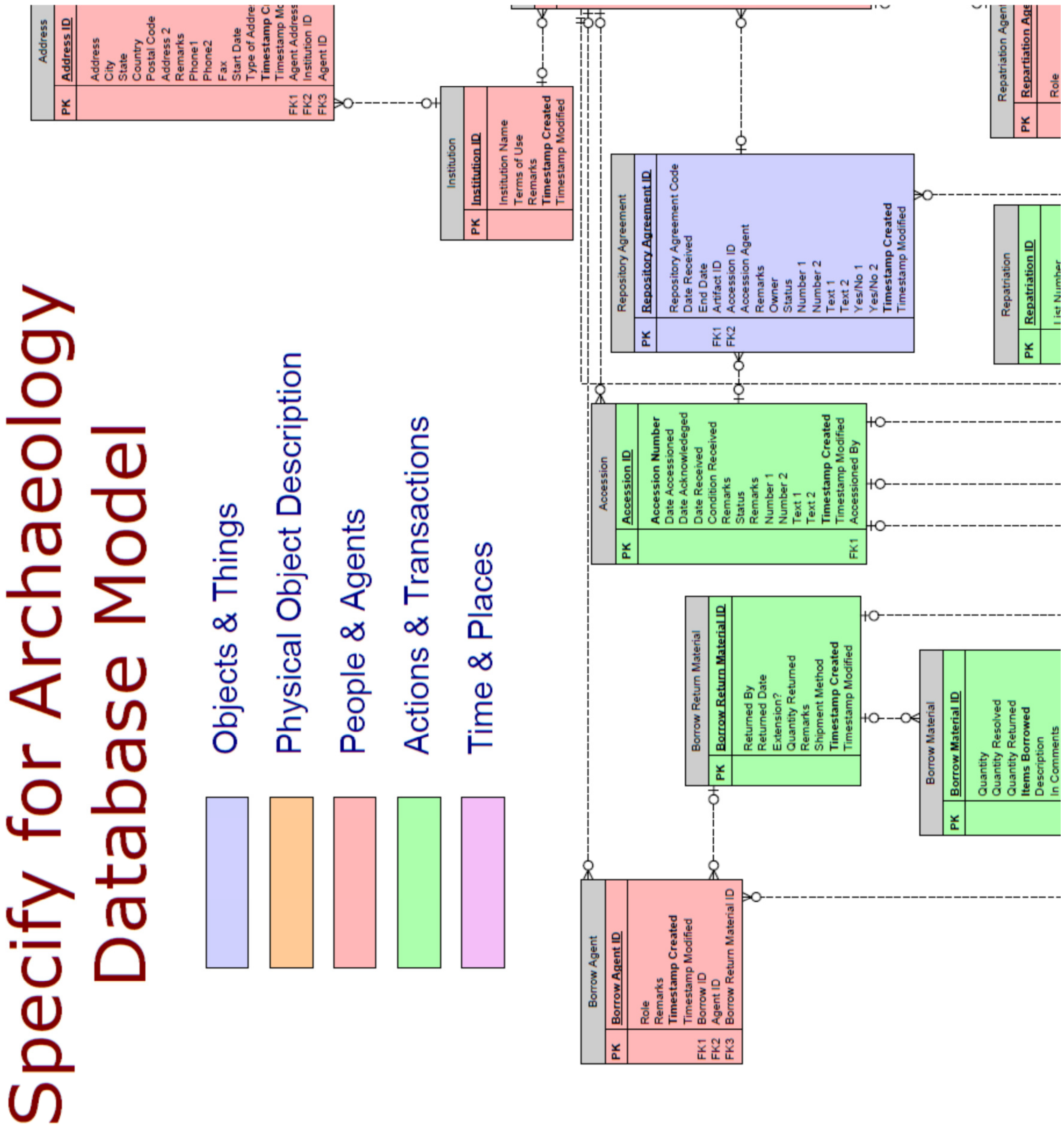


Figure 37 Section 2 of Data Model

Figure 38 Section 3 of Data Model

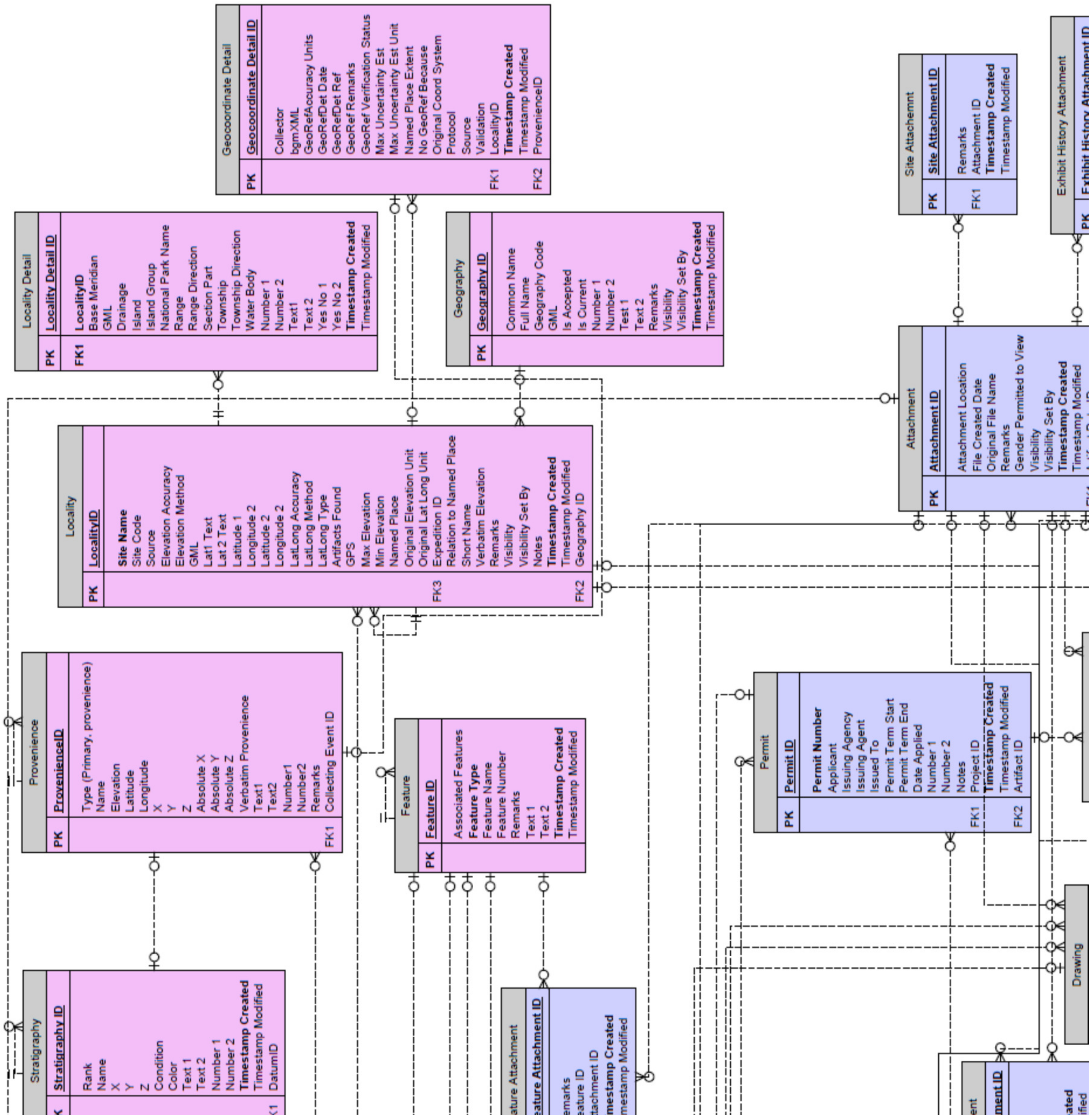


Figure 39 Section 4 of Data Model

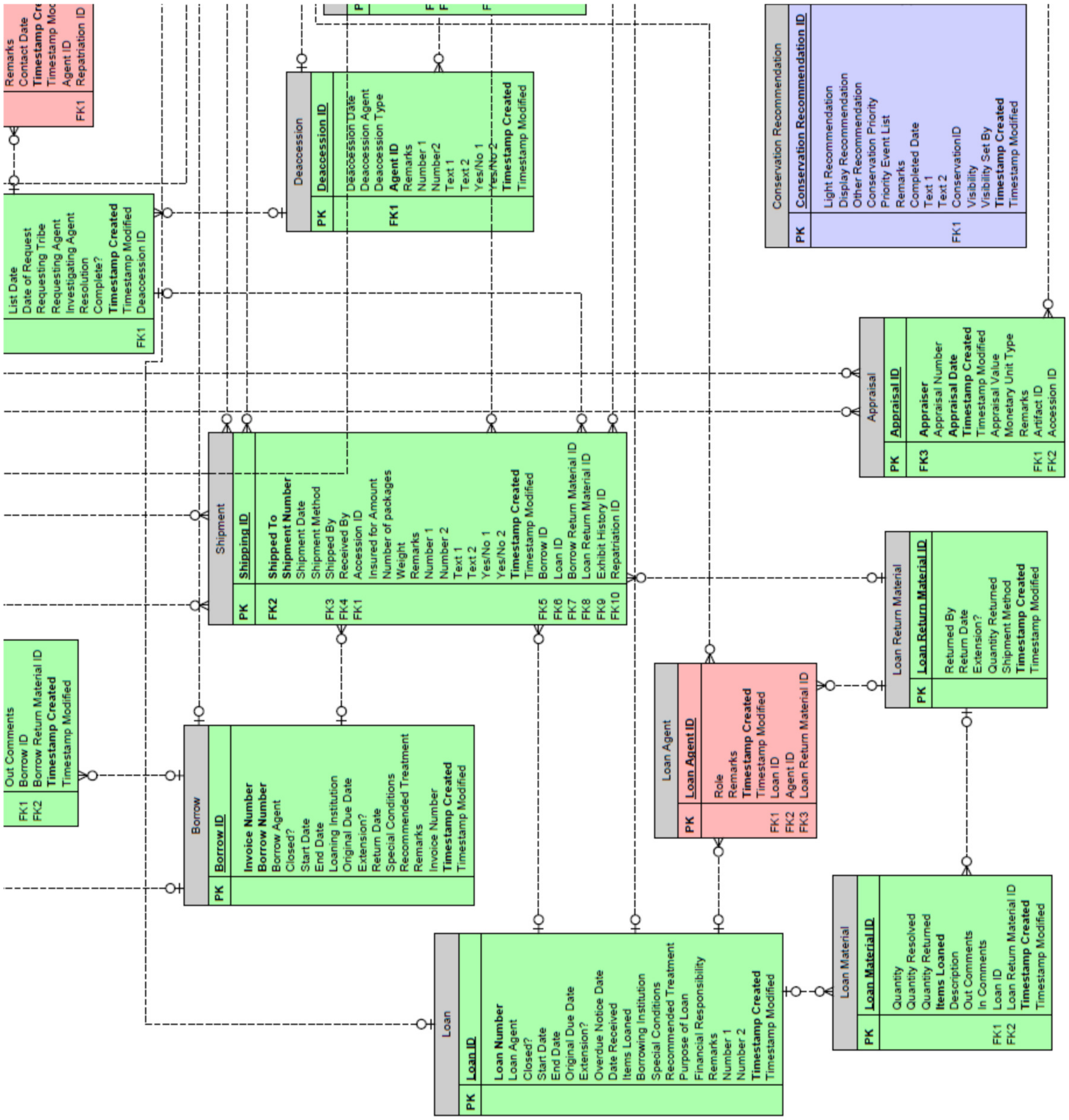


Figure 40 Section 5 of Data Model

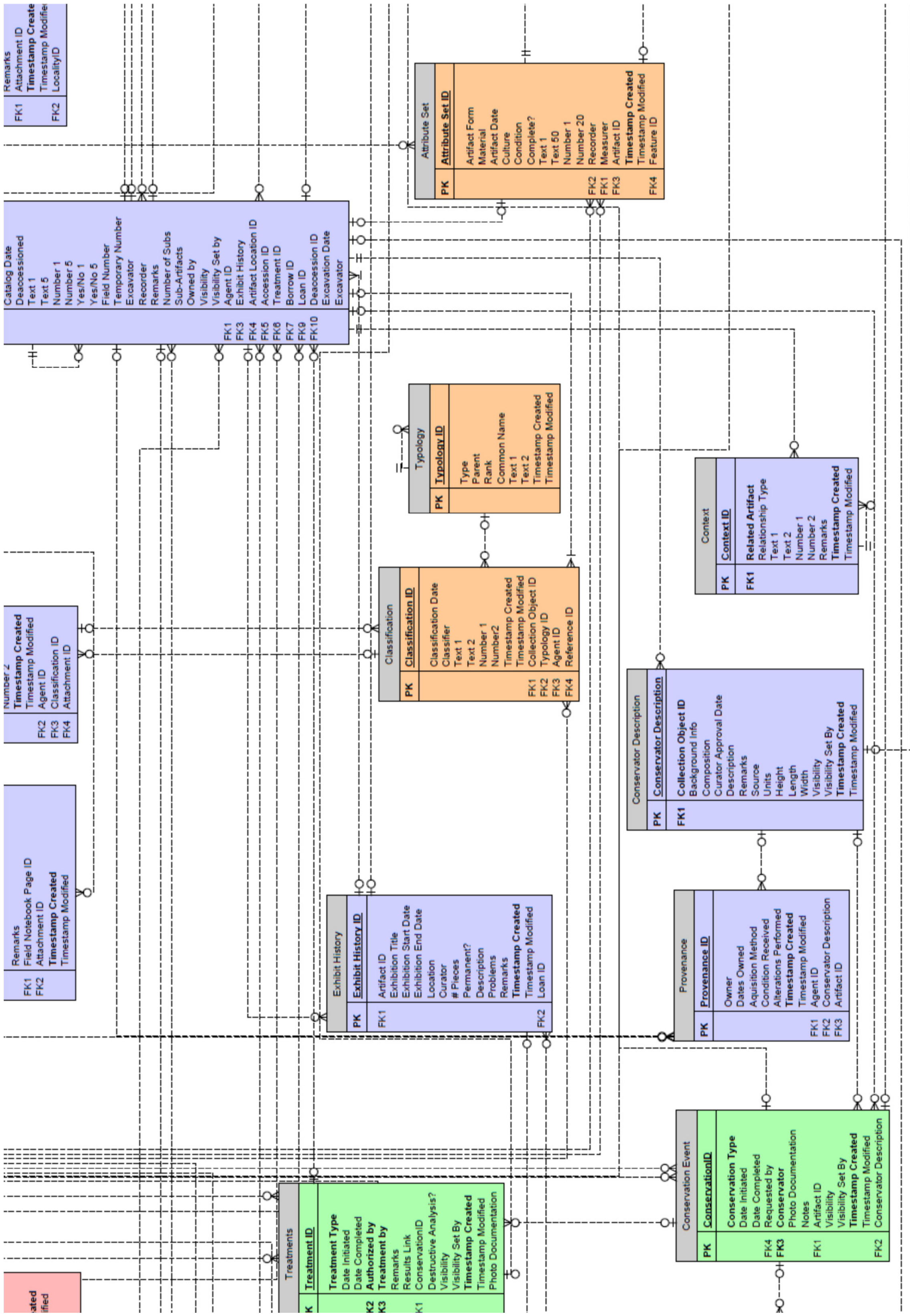


Figure 41 Section 6 of Data Model