FINAL NARRATIVE REPORT

PLANNING FOR THE LIFECYCLE MANAGEMENT AND LONG-TERM PRESERVATION OF RESEARCH DATA: A FEDERATED APPROACH

August 1, 2014

IMLS LG-51-12-0695-12

UNIVERSITY OF KANSAS LIBRARIES
GREAT PLAINS NETWORK
GREATER WESTERN LIBRARY ALLIANCE
INSTITUTE OF MUSEUM AND LIBRARY SERVICES
NARRATIVE

**Project Title:** PLANNING FOR THE LIFECYCLE MANAGEMENT AND LONG-TERM PRESERVATION OF RESEARCH DATA: A FEDERATED APPROACH

**Description of Project Partners:**
List the names and describe the contributions of any project partners as identified in your proposal.

1) *University of Kansas (KU Libraries).* The University of Kansas (KU) is a major research institution. KU served as the grant recipient and administrator. The primary investigator from KU, Deborah Ludwig, was responsible for administering the grant and for working with the project partners, the steering committee, and the program coordinator to meet the outcomes for the grant.

2) *Great Plains Network (GPN).* GPN’s purpose is “To support the research and education missions of our members through the use of advanced cyberinfrastructure. GPN actively supports the development of and provides support to research and education initiatives among the membership that extend or cyberinfrastructure in novel or unique ways.” GPN has supported this project through an in-kind match of 5% of co-investigator Dr. Greg Monaco’s time. Monaco is the Director for Research & Cyberinfrastructure Initiatives for GPN. He served on the steering committee and advisory council and was responsible for steering the discussion of infrastructure and took part in discussion of education, outreach, and training. Monaco provided significant contributions in planning the advisory council and data summit workshop events.

3) *Greater Western Library Alliance (GWLA).* The mission of GWLA includes delivering “cost-effective and high-quality information services and resources to its member institutions and their clientele [as a] a full partner and dynamic leader in the national and international scholarly arenas of learning and research [and] to capitalize on new technologies, to forge effective and meaningful partnerships, and to promote innovation and excellence.” GWLA has provided an in-kind match of 5% of Dr. Joni Blake’s time as co-investigator. Blake served on the steering committee and advisory council and provided significant contributions in planning the advisory council and data summit workshop events, in making connection to other national data management efforts, and in helping to discuss and disseminate the plan to GWLA member library deans.

Together the partners, GWLA and GPN, will lead the post-grant next steps.

**Overview:**
Provide a very brief (150 words or less) overview describing the need, problem, or challenge addressed by your project, the audience served, your project design, your intended results, and the extent to which you achieved your intended results.

Global-scale research drives creation of massive amounts of digital research data, most of which cannot be recreated if lost. Data has significant value when shared with other researchers over time. The DataFOUR project (*Data Federation Of University Research, or D4*) sought to understand issues involved in research data management and to engage regional stakeholders in developing a plan for collaborative solutions. The project design included: 1) an environmental scan; 2) workshops to bring together leaders and professionals representing research, library and technology interests from GPN and GWLA member institutions; and 3) creation and dissemination of a plan for multi-institutional data management for regional GPN and GWLA member institutions to guide post-grant efforts. Over the course of planning, emphasis shifted from the initial perceived need for

---

IMLS LG-51-12-0695-12
an immediate technical focus to address needs of stakeholders for education and training. Post-grant work will further develop plans for a pilot data repository project.

**Changes:**

*Provide a bulleted list of any changes in key personnel, budget allocation, scope, or schedule and reference the request to and approval by IMLS.*

- Project Coordinator, Barbara Paschke, named in the original project, was unavailable. Nicole Potter was hired for this role. Initial inquiry to Chuck Thomas was made on November 7, 2012 regarding how to proceed. Request to hire was sent to Chuck Thomas and Anthony Smith on January 25, 2013 and approved.
- For the first meeting of the advisory council in 2013, our originally named consultant was not available. We sent Kathy Mitchell biographies of two facilitators from KU (Dr. Marilu Goodyear and Dr. Noel Rasor) to help guide our planning efforts on June 5, 2013 via email and Ms. Mitchell responded that this was approved. We used the originally named consultant, Dr. James Mullins, in a second meeting of the advisory council in 2014.
- An approved one-year no-cost extension to the grant was approved in September of 2013. A new budget was submitted for the second year of the grant.
- An updated budget for year 2 was submitted to Dr. Sandra Toro on May 22, 2014 via email. The budget did not change the project scope, but made minor reallocations and corrected an error in calculating our in-kind salary contributions for the project. Not all funds allocated by IMLS were spent; however all local in kind contributions were met.

**Activities Completed During the Project:**

*List the high-level activities completed during the project and compare them to the activities you set out to complete for the project. Whenever possible describe activities in both quantitative and qualitative terms. If you did not complete all activities, explain why.*

**Grant Goal #1:** Undertake an in-depth environmental scan focused on current national and international data management initiatives and on the needs of our member universities for research data management services and infrastructure. [http://hdl.handle.net/1808/14857].

Lead by the PI and steering committee, volunteers from GWLA and GPN institutions completed all activities for goal 1 as specified in the grant application and met the stated success indicators. The environmental scan, shared with the advisory council in May of 2013, included:

- an in-depth literature review as part of the environmental scan;
- a survey of member institutions data management strategies also part of the environmental scan;
- a series of technical reports on approaches to research data management as part of the environmental scan;
- phone interviews, in lieu of site visits, with the University of Oregon, Arizona State University, University of Missouri, Texas A&M, University of Kansas, University of Oklahoma, Wichita State University, and University of New Mexico to better understand technical infrastructure and needs.

**Grant Goal #2:** Bring together a GPN and GWLA member forum and two-day workshop for the university research, library, and technology communities focused on understanding challenges and solutions in managing, sharing, and preserving research data. [Complete/exceeded]
• organized and facilitated two two-day workshops (2013 and 2014) and forum for representatives and our advisory council of leaders from member institutions as outlined in the grant proposal;
• developed an email forum for ongoing discussion and distribution of the outcome of the meeting and for future opportunities for involvement.

Grant Goal #3: Create and disseminate a plan for a scalable multi-institutional approach to research data management to support the university members of GPN and GWLA and advance this plan for funding. [Complete http://hdl.handle.net/1808/14857]

• completed and shared a draft plan and distributed the plan to the advisory council for comment via email in August 2013;
• completed and shared a final plan and an extracted “opportunities” list with the advisory council in May of 2014;
• shared the plan and subsequent findings with our grant consultant, Jim Mullins, and engaged in further discussion of steps toward strengthening shared infrastructure.
• supported GWLA member Colorado State University’s proposal to IMLS for the DataQ project as a resource for institutions and librarians supporting research data management. This project is tied to the first strategy of the plan, outreach, education, and training;
• developed a post-grant strategy and invited a small group of library leaders to discuss how to involve chief information and research officers in a pilot data repository project. This activity is directly tied to the second strategy of our plan, strengthening infrastructure.
• successfully pursued connections to the SHARE project and to ACRL’s interest in data management education and training for librarians;
• deposited the plan in KU ScholarWorks along with other grant materials.

We did not accomplish the activity of proposing “Recommendations for managing the governance and organizational dynamics necessary to sustain an inter-institutional model.” Within the timeframe of the grant, we found multiple initiatives that host institutions wished to partner to undertake with GWLA and GPN for the greater membership. In essence, GWLA and GPN are the governance and organizational entities for this work moving forward.

Project Results:
Analyze your entire project. Compare the actual accomplishments of the project with the intended results, outcomes, and impacts you identified in your proposal. Use quantitative data as well as qualitative examples, highlights from your evaluation, and compelling anecdotes. Identify and document significant project achievements and their value. Describe any significant unanticipated costs or circumstances that created obstacles to project success, and summarize lessons learned during the course of the project.

The project goals were to scan the environment to better understand data management challenges and the status of services at GWLA and GPN member institutions, to bring together leaders and stakeholders to discuss opportunities for collaborative action, and to put forward a plan for funding to member institutions for collaborative research data management.

Environmental Scan (Grant Goal #1). The activities for the environmental scan were accomplished. As part of the environmental scan, our first component consisted of a literature review conducted by a group of 11 volunteer experts from GWLA and GPN member institutions. The review was very useful in broadening our shared knowledge and the grants advisory council noted that the review was helpful for gaining bigger picture knowledge of research data
management issues and best practices. Advisory council members received the environmental scan in advance of the May 2013 meeting as preparation for meeting face-to-face.

The first component in the environmental scan was a literature review. Topics included: Introduction and general works; Assessment of researcher behavior, attitudes and needs; [Data] services, roles and responsibilities; Sharing, reuse, publication and citation; Data management planning; Policies and standards; Institutional repositories, approaches, and issues; Disciplinary or subject repositories, approaches and issues; Federated approaches; Economics/costs of data curation; Archiving and preservation; and Metadata and data description.

A second component of the environmental scan was a survey of GWLA and GPN member institutions. A group of volunteers from member institutions designed and conducted the survey, with the goal of better understanding services, policies, practices, unit roles, and infrastructure related to research data at member institutions. Forty-seven unique universities were sent an invitation to participate. The invitation was directed to three leaders from each institution: the chief officer for research, the chief information officer, and the university library dean/director. Response rate to the survey was low with only 38% of institutions responding. While we had hoped for a higher response rate, the outcome was useful information, particularly about the availability of infrastructure in the form of storage and repositories.

We learned, for example, that the service least commonly provided by respondents was “a repository for sharing data with appropriate access controls.” The most commonly provided service by central IT was “short term networked data storage” with long-term data storage (enduring more than 5 years) being less common. The service most commonly offered by the university library was “assistance with identifying national or international data repositories for archiving research.” With respect to institutional data policy, we learned that 70% of responding institutions have some general policy regarding data ownership, but fewer (42%) have policies that cover data once eternal funding has expired and less than half have policies covering research data ownership for data not supported by external sources of funding.
We also received valuable insights through an open-ended question in the survey, “What are the main institutional challenges for working with research data that your organization has identified?” Responses included issues of governance and leadership, staff expertise, funding, and faculty receptivity to data deposit:

- “Lack of centralized decision making and policies [everyone thinks it’s a good idea, but no one wants to take ownership]”
- “funding and staff expertise”
- “Not enough personnel to handle all of it. Lack of knowledge and subject expertise with the actually data.”
- “Getting faculty to deposit data sets and associated metadata”
- “…longterm continuation of services for unfunded [or expired funding] research”
- “Funding for long term storage/archive. Bridge funding for data storage in the mid term or between grants.
- “scaling services and support”

The third component of our environmental scan was the creation of a set of technical reports by volunteers from member institutions. Reports helped members become more familiar with technology and infrastructure for data management broaden knowledge about various technical approaches; however, this was probably the least impactful aspect of the environmental scan because technology changes rapidly and because advisory council of institutional leaders were less focused on technical strategies than anticipated as part of the planning process.

All together 23 people from GWLA and GPN took part in the environmental scan project. While there was certainly a strong impact in the knowledge we gained, a second enduring impact has been the development of community across different functional domains (IT, Library, Research) as people met and worked together on the environmental scan.

Success indicators for the environmental scan (grant goal #1) were met as outlined in our proposal. The environmental report was completed, reviewed, and subsequently used to inform and engage in dialog and planning. Iterative review by the Steering and Advisory Committees did reflect a successful articulation of the data management environment. This report may be found on our project website and archived in KU ScholarWorks. [http://hdl.handle.net/1808/14857](http://hdl.handle.net/1808/14857)

**Events. (Advisory Council Meetings & Data Summits, Grant Goal #2)**

The grant steering committee planned, organized and facilitated two workshops, the Big Data Summit in May 2013, and a second Data Summit 2 in May 2014. Each was held in conjunction with GPN’s annual meeting, which is co-sponsored annually by GWLA. The audience for the event was leaders and staff from GPN and GWLA member universities. This was one of the most successful outcomes of our grant.

Prior to our first workshop in 2013, we formed the grant advisory council of university leaders from libraries, IT and research. Dr. Marilu Goodyear, former library dean and former Chief Information
Officer at the University of Kansas, and Dr. Noel Rasor, assistant director of KU’s Public Management Center, facilitated two sessions of the advisory council during the 2013 summit. 32 people took part in the meetings. We used a series of techniques to help participants learn more about one another’s institutions and perspectives on management and services for research data. In addition to the advisory council meeting, we also offered a keynote speaker and opportunities to provide input into ideas for regional research data management initiatives by a broader group of participants. 135 persons attended the Big Data Summit, surpassing our original attendance projections.

Goal for the advisory council sessions were reviewed with participants:
• Participants engage in a meaningful discussion about their local institutions and what they aspire to do in data management
• Participants identify a range of options for collaborative services for data management
• Participants find common ground on one or two approaches they would like to consider further
• Capturing and funneling of the major directions to the planning team
• Going forward, participants remain engaged through some iterative design process that leads to the plan

Advisory council members subsequently worked together to identify areas of feasibility and criticality. The most feasible areas for collaborative action determined as the argument (the rationale for collaborative action), support for researcher data management plans, and education and training.

![Figure 1. Advisory Council Outcomes, Critical & Feasible Issues](image)

Funding models, business policy (and costs), and faculty incentives for deposit were considered critical by our advisors but less actionable at this time. Technical infrastructure issues of data storage and preservation fell out as less feasible and less critical. Discussion and follow up lead to a
plan with 5 initial strategies that was then narrowed to three strategies through email discussion during the summer of 2013.

Given the success of the 2013 data summit and the need for continued input from advisors, the steering committee held a second summit in May of 2014, bringing together 38 regional university leaders for advisory council discussion. Our consultant, Dr. James Mullins, Dean of Libraries at Purdue, facilitated the advisory council meeting and gave a keynote speech to a larger group of workshop attendees after the working meeting of the council. We saw an increase in participants at our second summit with over 140 attendees. Presentations from the 2014 workshop are available at: http://www.greatplains.net/display/Home/Big+Data+Summit+2014. Together, the impact from the two data summits and the work of the advisory council has been creation of regional community across functional domains of IT, Libraries, and Research. GPN and GWLA intend to continue offering these summits annually post-grant.

Success indicators for goal 2 were defined as completion of the advisory council meeting and workshop and positive response to an evaluation. 135 participants participated in the first data summit and over 140 participated in the second event in 2014. We did not receive sufficient response to the evaluation to meet the hoped-for evaluative criteria, but believe that increased participation in year 2 and the expressed desire of member institutions to see this event continue post-grant affirms its effectiveness and the ongoing interest in finding collaboration opportunities.

Create and Disseminate a Plan (grant goal #3)

Based on our work in conducting the environmental scan, meeting with our advisory council and attendees at the data summits, and follow up discussion by email with our advisors, we identified several key directions to serve as the foundation of a plan as part of the final goal and the overall outcome of this grant. Three strategies were adopted as the foundation of a regional plan:

1. **Strategy #1 – Outreach, Education, Training.** Offer a collaborative program of outreach, education, and training for member universities, addressing common institutional needs related to data management, data policy development, and best practices for data lifecycle curation.

2. **Strategy #2 – Strengthen Infrastructure.** Strengthen networked repository and storage infrastructure for data derived from research conducted by member institutions. Improve interoperability and opportunity between institutions, disciplinary organizations, and other research data hosts for working collaboratively with data and for access, archiving, and long-term preservation.

3. **Strategy #3 – Communication and Advocacy.** Develop and execute a strategic communications plan and program of advocacy to advance member initiatives in data management and to advocate for data sharing as part of a greater ecology of scholarship.

Under the leadership of steering committee members, we again formed teams of volunteers from member institutions to flesh out the actions deemed representative of institutional needs for strategies #1 and #2. To identify opportunities under strategy #1, volunteers worked together to identify education and training workshops that GWLA and GPN might sponsor and worked with Colorado State University, a GWLA institution to support DataQ, a proposal to create an online collaborative space for sharing research data questions and answers, crowdsourcing and curating
responses to research data questions, identifying resources to assist with research data questions, and establishing best practices for answering research data questions. For strategy #2, we conducted phone interviews with leading institutions that provide relevant existing and emerging management models. The phone interviews consisted of eight member institutions including University of Oregon, Arizona State University, University of Missouri, Texas A&M, University of Kansas, University of Oklahoma, Wichita State University, and University of New Mexico. Strategy #3 was addressed by the steering committee through contacts with ARL, the SHARE project, ACRL, and through development of a robust social media presence. We have been involved in advocacy for institutional and library roles associated with the OSTP memo and subsequent SHARE proposal from ARL, AAU, and APLU by writing letters and signing calls to action. We have email lists of over 150 people for general RDM announcements. We have named our project D4 (Data Federation Of University Research). We have a Facebook presence and a Twitter presence with 116 followers that rapidly increases.

Success indicators for goal 3 included the development of a regional plan with three strategies and activities for ongoing work post-grant. We met the goal of creating a plan and the plan is driving regional collaboration in education, outreach and training; strengthening infrastructure, and advocacy. While some activities under the strategy have been completed or are ongoing, others provided the bridge to continue the D4 project vision as a living and growing collaborative effort for GPN, GWLA and their member institutions.

**Unanticipated outcomes**
An unanticipated outcome for the grant steering committee was the direction our advisory council set. As we were developing and submitting the original grant proposal, we envisioned a project much more focused on scalable multi-institutional infrastructure. Over the course of the planning process, we discovered the strongest desire for collaboration among member institutions centered on education, outreach, and training. Our planning efforts shifted our roadmap towards constructing a network of people, skills and knowledge. In part, the shift away from infrastructure might be a result of the Office of Science and Technology Policy memo, *Expanding Public Access to the Results of Federally Funded Research* (February 22, 2013), which created both excitement and uncertainty regarding technical solutions for managing research data. We anticipate post-grant work to further consider opportunities to pilot an infrastructure project or projects.

**Summary Outcomes and Impact**
In terms of overall outcomes, our efforts in education and outreach have been notable. We have created a sense of community within the GWLA/GPN region for professionals who manage research data. Through this project, we have involved more than 200 people representing GWLA and GPN member institutions. This number includes individuals who volunteered for various working groups, members of the advisory council, and professionals from higher education who attended the advisory council meetings and data summits in 2013 and 2014 to engage in meaningful discussion about issues surrounding management of research data and the state of data management at our regional institutions. In 2013, we settled on the name for our project DataFOUR (or D4, the Data Federation of University Research) and created a robust social media presence that continues to flourish. Our interests are now formally connected to the Association of Research Libraries’ SHARE project with the appointment of the GWLA’s Joni Blake to one of the SHARE project teams. We have also pursued discussion with the Association for College and Research Libraries
(ACRL) about a traveling roadshow approach to institutional education. GWLA and GPN are working together to sponsor new regional data management workshops (see What’s Next? below) in the coming year. Both membership organizations, GPN and GWLA, now have a more diverse set of professional partners and expanded perspectives and knowledge for all parties involved. A community of professionals has emerged from different functional domains to plan and work together.

For GPN, this grant has increased the partnership and communication among the technology, library and science communities, which has now become an objective in the GPN Strategic Plan. Members of GPN have expanded their skillsets and understanding, surpassing the traditional technology approach. They have carved out a niche to incorporate topics that were once deemed in the library only domain. For instance, Research Data Management has become a topic covered in the GPN Professional Development Program (a series of webinars with a national audience that are held approximately eight times per year).

GPN’s Annual meeting, with the development of the Data Summit workshops, is now a joint meeting with GWLA and attracts a more diverse representation from technology, library science, high performance computing, and research data management from member universities. GPN now selects keynote speakers for their ability to bridge multiple areas of interest in hopes that it will change the attitudes and views of other individuals in this community. GPN has been asked by national organizations (e.g., The Quilt) to discuss the project and partnership as a model for other regional networking organizations.

As the second key partner in the D4 project, the Greater Western Library Alliance has also experienced significant impacts throughout these past two years. The executive director, Joni Blake, reports that GWLA’s participation with the Great Plains Network (GPN) and KU Libraries in the IMLS grant has been very rewarding and allowed the organization to move a significant element of GWLA’s strategic plan forward. By working more closely with these partner organizations, GWLA has been able to engage a new group of librarians within the community of GWLA member institutions.

Blake notes, “In addition to the expected outcome of investigating the feasibility of a multi-institutional, multi-consortial solution for research data management, we are delighted that several unexpected outcomes also emerged. These outcomes include joint-sponsorship of the GPN Annual CyberInfrastructure Conference, sponsorship of multiple E-Science Bootcamp Workshops, and we are planning to sponsor a Software Carpentry Workshop in the fall of 2014.”

What’s Next?

After receiving positive feedback from our grant’s activities, we have identified potential projects to build on the grant’s success. KU Libraries, which took the lead during the grant period, will return to the role of member institution. GWLA and GPN will assume responsibility for ongoing initiatives moving forward. Joni Blake, GWLA, and Greg Monaco, GPN, will provide leadership to the steering committee and continue to support and incubate ongoing projects. Near term initiatives anticipated post-grant include:
• DataQ – a collaborative project lead by the University of Colorado with GWLA and GPN as sponsoring partners.
• Potential partnership with the ACRL in a roadshow or workshops for librarians on research data management.
• Regional software carpentry workshops for librarians and data management specialists.
• Development of a graduate student training program online in research data management based on training efforts already in place at GWLA member institutions.
• Development of a pilot project for a regional repository for research data. Initial conversation will include 6 institutions from the GPN to begin with a 4-state pilot. (University of Arkansas, University of Missouri, University of Oklahoma, Oklahoma State University, University of Kansas, and Kansas State University.)
• Ongoing Data Summits in connection with the Great Plains Network annual meeting, co-sponsored by the Greater Western Library Alliance.
Grant Products.

1. Environmental Scan. The initial (preliminary) report to the advisory council on the environmental scan is housed on the project web site and permanently archived in KU ScholarWorks. These may be downloaded as the full document or individual components (literature review, technical reports, survey) are available for individual download. http://kuscholarworks.ku.edu/dspace/handle/1808/14857

2. MULTI-INSTITUTIONAL DATA MANAGEMENT: A PLAN

3. 2014 Data Summit Presentations http://www.greatplains.net/display/Home/Big+Data+Summit+2014

4. Project Web Site. In order to create awareness and organization for our project, project staff created and maintained a project website, which included all of the products created during this grant. (http://imls.gwla.org)
5. Social Media:

**Twitter:** @DataFOUR

**Facebook:** [www.facebook.com/D4DataFOUR](http://www.facebook.com/D4DataFOUR)