

Interim Performance Report Narrative

October 1, 2012 – September 30, 2013

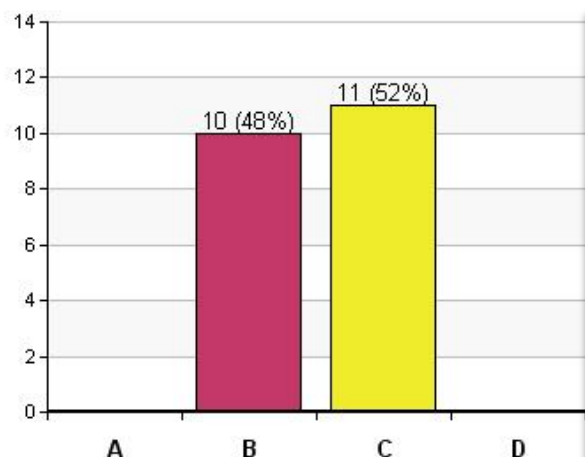
Deborah Ludwig, Principal Investigator

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

The purpose of this IMLS National Leadership Planning grant is to propose shared ways of approaching the problems facing institutions with respect to management of researchers' data. More specifically:

The University of Kansas (KU) Libraries in collaboration with two partners, the Greater Western Library Alliance (GWLA) and the Great Plains Network (GPN), seek a one year (Oct 2012-Sep 2013) IMLS National Leadership Grant designed to leverage collective strengths and create a proposal for a scalable and federated approach to the lifecycle management of research data based on the needs of GPN and GWLA member institutions.

A shared approach to the issues associated with research data resonates with members of this grants' advisory council who are primarily CIO's, Dean's of Libraries, or other University leaders involved in university research initiatives. The chart below shows the outcome of discussion during a meeting of the advisory council held on May 29, 2013.¹



- I believe that research data issues should be addressed by:
 - A. Each individual **campus** alone.
 - B. Each **individual campus with assistance** from national and regional groups.
 - C. **Regional groups should provide services** to collaborating campuses, with each campus implementing the services they see as important.
 - D. **National groups** should take full responsibility.

¹ The use of voting "clickers" allowed council members to anonymously voice beliefs and positions throughout our meeting.

Goals of the Grant and Accomplishments to Date

These three goals were established for this grant. Goals 1 and 2 were completed during the first year of the grant. Goal 3 began in the first year of the grant and we continue to work on this goal as part of a no-cost grant extension approved for a second year of work.

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.

Goal #2: Bring together a GPN and GWLA member forum and two-day workshop for the university research, library, and technology communities focused on challenges and solutions in managing, sharing, and preserving research data.

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.

People undertaking the work of this grant include a steering committee of leaders in libraries and cyberinfrastructure² working with an advisory council of university leaders³. A large amount of research and planning work to date has been accomplished by the working groups established for the environmental scan and by the current working groups who are developing a road map of projects connected to the strategies in our emerging plan.

Accomplishments for Goal 1, Environmental Scan (Spring 2013)

Our first goal, an in-depth environmental scan, is complete. The scan consisted of an extensive literature review, a survey of GWLA and GPN member institutions, and a series of informational reviews of enabling technologies and technology providers.

The literature review covered 12 areas: an introduction and general works related to the topic of research data management; literature assessing researcher behavior, attitudes and needs; data services, roles and responsibilities; sharing, reuse, publication and citation of research data; data management planning; policies and standards; institutional repositories; disciplinary or subject repositories; federated approaches to data management; economics/costs of data curation; archiving and preservation of research data; and metadata and data description. Experts from various member institutions covered each topic. The literature review was made available to the advisory council and can be downloaded from the project website at: <http://imls.gwla.org/file-cabinet/literature-review>. Once the project is complete, the literature review will be deposited permanently in the University of Kansas' institutional repository.

The survey of member institutions was also completed in the spring of 2013 and made available to the project's advisory council. The survey was sent to three contacts at each institution with potential to contribute to an understanding of each institution's services, policies, practices and infrastructure related to research data. The contacts selected were the chief administrative officers for the university libraries, information technology, and

² <http://imls.gwla.org/the-team/project-leadership>

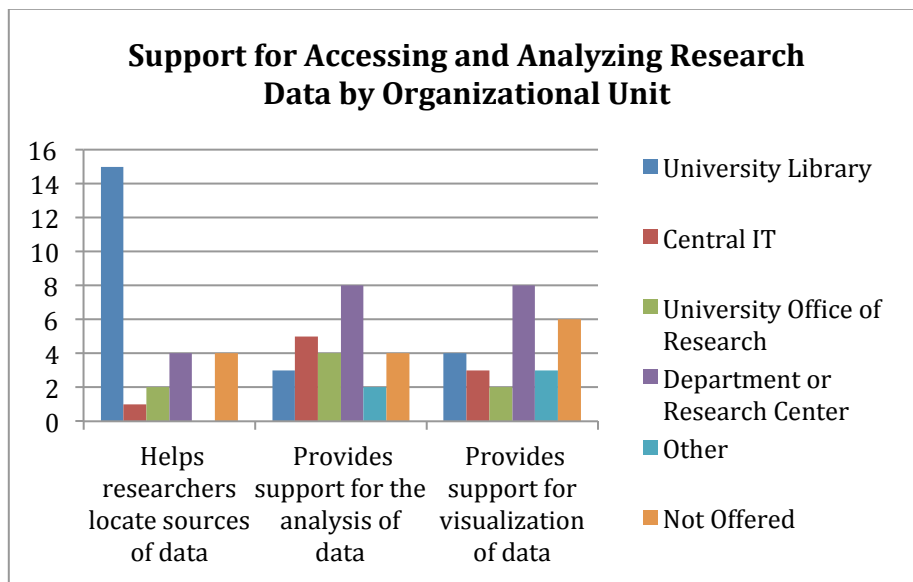
³ <http://imls.gwla.org/the-team/advisory-committee>

academic research. Institutions were invited to confer and to send either a single response on behalf of the institution or to send separate responses from multiple divisions (library, campus technology, research). A total of 123 surveys were sent to 47 institutions. 23 individuals from 20 institutions responded to the survey. Based on individual returns, the response rate was 18%. Based on institutional response, the response rate was 38%. A report of survey findings is available on the web site.⁴

In the section of the survey concerned with service provision, five main service categories were addressed:

1. General Support and Services for Research Data
2. Storage, Archiving, Preservation, and Sharing of Data
3. Accessing and Using Research Data
4. High Performance Computing
5. Support services management, preservation, and access to digital and non-digital research data for long-term access to campuses

The survey offered us a better understanding of what roles different university divisions play with respect to various data services within institutions. For example, the chart below shows how universities libraries most frequently assist researchers with finding data, while campus and research centers most frequently offer data analysis and visualization support.

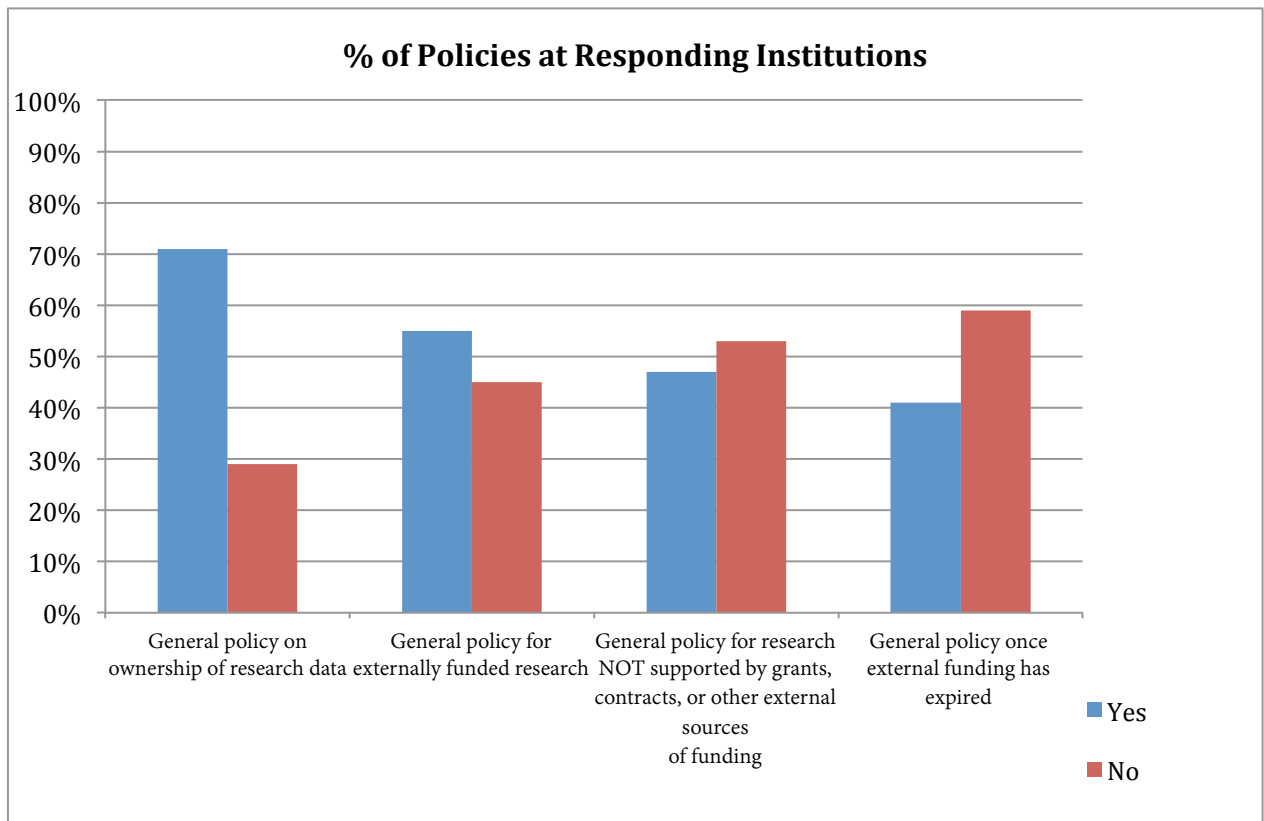


In addition to looking at various service areas related to the lifecycle of research data, the survey asked institutions about their institution's data policies in accordance with four possible policy types:

1. General research data ownership policy
2. General policy for externally funded research
3. General policy for research NOT supported by grants, contracts, or other external sources of funding
4. General policy for research data after external funding has expired.

⁴ <http://iims.gwla.org/file-cabinet/survey-of-institutions>

Not all survey participants answered our questions regarding research data policy. Questions regarding policy received 17 responses. 12 (71%) of respondents indicated that their institutions have a general data ownership policy. More specific policy areas had fewer affirmative responses. For example, fewer than 50% of respondents reported having policies for research NOT supported by grants, contracts, or other external sources of funding or policies that cover data once external funding has expired.



Finally, we asked respondents to identify the main institutional challenges for working with research data through a free text answer. Key challenges included:

- Funding for sustainability of efforts. Respondents noted the need for business model to assign costs as well as the need for funding to cover both service providers and researchers.
- Coordinated decisions, policies, and roles. As one respondent commented, “everyone thinks it’s a good idea, but no one wants to take ownership.”
- Awareness of the issues involved in the preservation and archiving of data.
- Lack of technology and infrastructure.
- Researcher attitudes regarding data and data sharing.
- Staff expertise in terms of not having sufficient personnel and lack of knowledge and subject expertise to work with the actual data.

The third component of the environmental scan consisted of a series of what we called “technical reviews.” Topics covered by these reviews included:

- Major Research Data Initiatives and Approaches to Curation
- Working Repository Examples (Institutional or Disciplinary)
- Repository Technologies
- Metadata Services & Schema

- Data Discovery / Access / Sharing
- Storage
- Preservation Services / Systems

The technical reports are also available through the project web site.⁵

Altogether, the work on the environmental scan involved 23 individuals from 12 different GWLA and GPN institutions and representing different areas of higher education, including information technology, research support, and libraries. We believe the diverse participation in this project has been a hallmark of its strength and has contributed to a key outcome specific to goal 1, but also important to the project in general: *a shared understanding of the scope and dimensions of international and national approaches to the problems of lifecycle management of research data along with the challenges and needs of researchers in our member institutions.* We were able to build on the outcomes of the environmental scan in undertaking our work in goal two, which involved bringing together our advisory council and facilitating a larger gathering of participants at a gathering dubbed simply, *The Big Data Summit.*

Accomplishments for Goal 2, Member Forum & Workshop (Spring / Summer 2013)

In the spring of 2013, we formed an advisory council comprised of university leaders from libraries, campus information technology, and institutional research support. Members⁶ of the advisory council represented the following GWLA and/or GPN institutions:

- Arizona State University
- University of Arkansas
- University of Colorado
- Iowa State University
- University of Kansas
- University of Kansas Medical Center
- Kansas State University
- University of Missouri
- University of Missouri, Kansas City
- University of New Mexico
- Oklahoma State University
- University of Oregon
- Texas A&M
- University of Tulsa
- University of Washington
- Wichita State University

We brought the advisory council together in meetings and at a workshop held in Kansas City on May 29-30, 2013. The meetings of the advisory council were lead by Marilu Goodyear, former library dean and former Chief Information Officer at the University of Kansas. 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.

⁵ <http://iims.gwla.org/file-cabinet/technology-infrastructure-reports>

⁶ <http://iims.gwla.org/the-team/advisory-committee>

As a result of our conversations, the advisory council identified issues and areas that they felt were either critical areas for action or areas that were most feasible. For example, in our first meeting, council members were asked to brainstorm issues from the perspective of their institution.

Identified Major issues Identified from School Reports By Consultants Goodyear & Rasor

- Who has ownership over this issue: Research, IT, or Libraries?
- Who has ownership over the data: IP ownership issue and confusion?
- Cost issues incentivize "wrong" researcher behavior. Models for faculty incentives, potentially inter-institutional. Also need for persuasive language and framing (relating to institutional and researcher goals) around the need for data mgmt.
- Training as multi-institutional possibility. Working with NIH and other funder requirements.
- Data management is not the same as Open Access. Need for access controls and curation processes to remove sensitive data, etc. Importance of de-identification and other business requirements. Challenge of privacy and need for heightened security.
- Models for shared understanding of need between Libraries and IT-- different focuses?
- Who are the collaborators and partners? Different opportunities and potentials for collaboration at university, system, regional levels.
- Underlying infrastructures will determine what collaborations are acceptable, feasible, and reasonable. Researcher has to view that the access of that data will not be hindered.
- Private sector motivations around ensuring compliance with new mandates
- Framing and communication with researchers.
- Barrier of decentralized funding models in universities.
- Lack of sustained administrative support.
- Need to develop institutional policies around data management and retention.
- Need for centralized repository-- but what kind?
- Influence of local organizational culture-- focus on compliance, access, etc.--and excellence determined by this environment.

From this wide-ranging set of issues, we clustered into a smaller set of issues in order to undertake some prioritization. The table below shows the outcome of one of the culminating activities, a “dot voting” exercise where participants determined which issue clusters they felt were either more critical or more feasible.

	Red - critical	Blue - feasible
The Argument	11	12

Funding Models	9	2
Business Requirements/ Policy	8	5
Faculty Incentives	8	0
Data Management Plans	3	9
Discovery and Access	2	0
Storage	1	1
Education and Training	0	7
Preservation	0	4

Out of a group of potential issues discussed by the advisory council and steering committee, a very broad issue deemed most critical and feasible was articulated summarily as developing “the argument” for change strategies to support research data management as a multidisciplinary issue with benefits for university research. Other issues identified were funding models and work to address business requirements and policy development. Not all issues identified as critical were considered highly feasible. Incentives for faculty to share data, for example, were subsequently discussed and recognized as an impractical target for action by this group even though this issue is considered relatively important.

Conversely, several issues were considered feasible, although voted less critical such as shared approaches to data management planning support and education and training. Issues that rated lowest in both criticality and feasibility were data discovery and access, data storage, and data preservation. This outcome was fairly surprising because issues related to people and skills, repositories, storage, and discovery did feature prominently in discussion of issues during our meetings and other assessments, but they did not emerge as areas of first concern for leaders in this round of conversation. To some extent, the focus on change strategies (“the argument”) may have also reflected the recent memorandum by the Office of Science and Technology Policy (OSTP) directing federal funding agencies to establish plans for sharing the outcomes of funded research.⁷

We also tapped into attendees at our *Big Data Summit* event held concurrent with the advisory council meetings to further explore institutional needs with respect to research data management. We used an exercise with the 135 participants at the Summit where we asked them to tell us the following:

- Name some things you would envision as elements of a good solution to issues of research data management at your institution.
- Name some stakeholders (people, positions or roles) that must be involved for efforts to succeed.

⁷ <http://www.whitehouse.gov/administration/eop/ostp/library/publicaccesspolicy>

- Please share any ideas you have on how institutions could cooperate on data management and curation.

We received 180 comments from the participants of this exercise. Elements that participants envisioned as elements of institutional research data solutions clustered into these categories: Skills/Expertise/Personnel; Service or Policy Definition; and Best Practice & Procedures.

Accomplishments for Goal 3, The Plan

Based on our work in conducting the environmental scan and the meetings held as part of goal 2 with the advisory council and attendees at the Big Data Summit, 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. The initial 5 strategies were:

- *advocate for the adoption of policies for consistent and federated approaches to managing and sharing data resulting from research sponsored by federal agencies*
- *develop a joint program of outreach and education for member universities to address institutional needs related to data management, including policy development, practices related to stewardship for data management, and basic levels of education and training*
- *identify opportunities for strengthening infrastructure in member institutions, improving interoperability between institutions and disciplinary organizations for sharing data, and develop new platforms to meet needs for data management that are not universally addressed at member institutions*
- *create a strategic communications program to build awareness about the GWLA/GPN data lifecycle project, to develop common talking points about both national and institutional initiatives that could be used by member institutions, and to spur collective action in response to emerging national and international developments related to research data.*
- *create an assessment strategy for institutional and overall project progress in developing policies, strategies, partnerships, and resources for research data management.*

Through subsequent vetting with the advisory council and among the members of the steering council, we reduced the plan to three key strategies. We are now working to develop a project map and final plans for vetting and to seek funding. The three strategies are listed below. A full report on the initial roadmap available on the project web site.⁸

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.

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,

⁸ <http://iimls.gwla.org/file-cabinet/multi-institutional-data-management-a-plan>

disciplinary organizations, and other research data hosts for working collaboratively with data and for access, archiving, and long-term preservation.

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.

Working groups made up of individuals from member institutions are now developing the projects associated with strategies 1 and strategy 2 and these, along with suggested funding models, will be made part of the final plan.

Under strategy 3, communications and advocacy, 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 for general announcements (over 150 recipients), for our working teams, advisory council, and steering council. We have named our project D4 (*Data Federation Of University Research*). We have a Facebook presence and a Twitter presence with a rapidly growing number of followers. We most recently presented on the D4 initiative at the Coalition for Networked Information Fall Meeting.⁹ As part of the final plan, we will formalize our ongoing communications plan.

Next Steps & Grant Completion

Within the steering committee we are discussing the final steps necessary to formalize the plan and its projects and to develop the overall buy-in and project endorsements we believe are necessary for success. We are considering a second face-to-face meeting of the advisory council for May 2014, modeled after our initial successful meeting in 2013. We are also discussing opportunities for the boards of GWLA and GPN to preview and discuss the anticipated plan in advance of that meeting.

Much has changed since 2010 when this project was first envisioned. Particularly in this last year, the OSTP memorandum of February 22nd and the emergence of the SHARE proposal have caused us, along with our advisors, to focus our priorities differently with perhaps less emphasis on the immediate creation of deep infrastructure and more emphasis on services and making data discoverable and shared. We see this initiative and the projects it will subsequently generate as contributors to a larger ecosystem of initiatives that should ideally fit together to advance knowledge among our multiple institutions.

⁹ <http://www.cni.org/topics/digital-curation/f13-wang-collaborating/>