

ABSTRACT

Title: Planning Research Data Services in Academic Libraries: Designing a Conceptual Services Model based on Patron Needs Assessment

This planning project will serve as an initial investigation for a larger project aimed at developing research data services in academic libraries to support researchers in data collection, management, analysis, presentation, and preservation. The goal of this project is to design a research data services model at the conceptual level based on a needs assessment of researchers in academic libraries as well as to suggest guidelines for library and information science (LIS) curricula for research data services. In this project, research data services refers to a range of library services to assist academic researchers to collect, manage, analyze, present, visualize, and distribute data in their research activities. For example, research data services can potentially include a wide variety of services, such as data collection, data reference services, data refinement, data storage and management, data analysis, big data analysis, data visualization, relevant workshops, and other services. The planning project proposed herein focuses on the perceived data-related needs of diverse patrons and stakeholders, such as library patrons, librarians, and LIS scholars. The four major objectives of this planning project are: (1) to understand current status of data services; (2) to assess researcher community needs; (3) to identify potential services feasible in the field; and (4) to suggest curricula for formal education of data services librarians. To achieve the research objectives, the project team will conduct case studies, surveys with potential patrons and academic librarians, interviews with LIS scholars, and focus groups with different stakeholders. The expected outcomes include: a list of potential data services feasible in operating libraries, situations in which patrons need research data services, resources needed to offer data services, knowledge and skills needed to data services librarians, curricula suggestions for data-related LIS programs, and others. In particular, a conceptual data services model will be produced, which will identify types of data services, associated resources necessary for services, service platforms, knowledge and skills needed by librarians, and corresponding librarian education plans. This conceptual model will be used in a subsequent project to develop a research data services prototype, which will be implemented, tested, and enhanced in an operating academic research library. The outcomes of this project will facilitate data services practice across research universities in the United States, and it will lead to a significant advancement of nation-wide research support capabilities in the emerging computational science environment.

Planning Research Data Services in Academic Libraries: Designing a Conceptual Services Model based on Patron Needs Assessment

STATEMENT OF NEED

Project Overview

This planning project will investigate: (1) current status of research data services; (2) researcher community needs assessment; and (3) perceptions and opinions of heterogeneous stakeholders. Based on the findings from this initial investigation, we will suggest a research data services model at the conceptual level. The *conceptual research data services model* will identify types of data services, associated resources necessary for services, service platforms, knowledge and skills needed by librarians, and data services librarian education plans. In addition, the project will suggest curricula for educating data services librarians in library and information science (LIS) programs. The conceptual model will serve as a framework for research data services development in academic libraries as well as librarian education in LIS programs.

Definitions

In this project, *research data services (RDS)* refers to a range of library services to assist academic researchers to collect, manage, analyze, present, visualize, and distribute data in their research activities. RDS potentially include a wide variety of services, such as: (a) data collection; (b) data reference services; (c) data refinement; (d) research data management; (e) data analysis; (f) big data analysis; (g) data presentation and visualization; (h) relevant workshops or tutorials; and other services.

Research data management is part of RDS that encompasses broad activities including data management planning, data collection and repository, digital preservation, and data literacy (Witt 2012). It can be described in terms of a system of people, policies, resources, and technology that support and give direction to researchers and organizations as they produce, collect, use, and preserve research data (Steeleworthy 2014).

A conceptual RDS model will be produced from this project. The *conceptual RDS model* will identify types of data services, associated resources necessary for services, service platforms, knowledge and skills needed by librarians, and training and education plans of data services librarians.

Emerging Needs of Research Data Services

In the research community, the paradigm has recently shifted to computational science across various disciplines. Computational science deals with large, complex datasets that require advanced data management techniques. Whereas traditional empirical studies utilize well-organized data mostly coming from controlled environments, computational studies involve large-scale analysis and simulation of enormous, unstructured data collections. The nature of research data has changed from predominately “structured” to increasingly “unstructured” data (e.g., raw text, raw numerical data, images, and/or semi-structured web data) as sources of data have diversified to web, digital documents, transaction logs, OCR, RFID, and sensors. Data analysis techniques have rapidly evolved with the exponential increase in computational power, and data science has emerged as a new tool to manage and analyze the resulting large, complex, unstructured datasets.

The emergence of data science as a mechanism for managing complex, computational research has motivated academic libraries to develop RDS for their faculty and students. Funding agencies increasingly require data management and sharing plans, which is forcing researchers to devise plans for data management and sharing to compete for external funding (Fearon & Association of Research Libraries, 2013; Tenopir *et al.*, 2014). To meet this need, academic libraries are developing infrastructure and services to support researchers in establishing and implementing their data management plans (DMP). Most of existing services, however, focus on assisting with the development of the DMP and managing the data products of research projects, e.g., data preservation and archiving. This project not only investigates the need of patrons in relation to data services but also explores opportunities to expand RDS to more directly engage the library in the research process, such as data collection, documentation, analysis, and other services.

Need for New Roles of Academic Libraries

Academic librarians traditionally support researchers' access to prior literature through collection and reference services, and assist with the dissemination of scholarly output at the end of the research process through various channels of scholarly communication. As research environments change, however, new expectations are emerging in librarianship (O'Malley, 2014). Librarians are increasingly expected to engage directly and indirectly with researchers throughout the research process (Swanson & Rinehart 2016). In this new research environment, librarians are providing support locating and accessing data collections, as well as storing and disseminating data products (Bracke 2011). Data analysis and presentation can be also potential areas that directly engage librarians in the research process. As shown in Figure 1, data services can expand library services to align more closely with the research process, including data collection, storage, analysis, result presentation, and data curation and distribution.

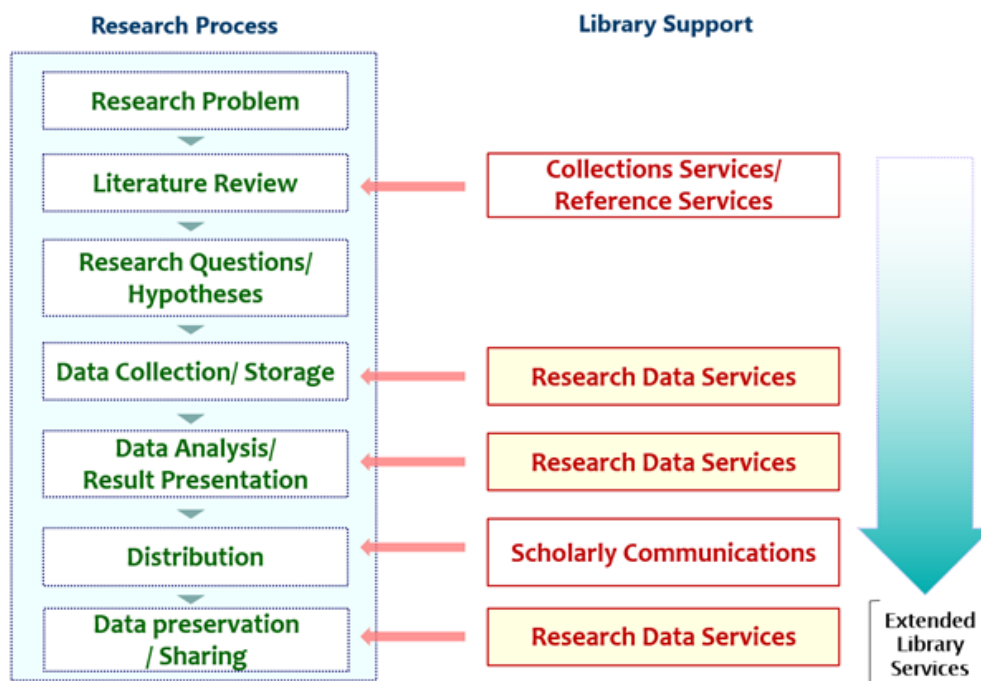


Figure 1. Research process and corresponding library services

Need for Educating RDS Librarians

In recent years, new titles of librarianship have appeared in relation to RDS, including data management librarian, e-science librarian, data visualization librarian, and data services librarian. Despite the increase in data-related positions, however, the necessary skills needed to succeed in these new roles remain poorly defined. Many academic libraries are searching for ways to equip their librarians with the skills necessary to offer these new services. Similarly, library and information science (LIS) programs recognize the need for curricula to prepare a new generation of data-savvy librarians for data services roles. Lack of knowledge and skills among library staff has been a challenge in development of specialized data services (Corrall, Keenan, and Afzal, 2013). Therefore, education and training is important to realize functional RDS in academic libraries. This project will investigate the skills needed for data services librarians as well as relevant subjects that can be covered in LIS education.

Limitations of Existing Research Data Services

When the National Science Foundation (NSF) announced that data management plans would be required along with grant proposals beginning in 2011, academic libraries in research universities increasingly began to provide data services, so called data management planning (DMP) services. DMP has focused on "data management," which concerns how to organize, store, preserve, and share the research data produced as one of the products from the project (Peters & Dryden, 2011; Tenopir, Birch, & Allard, 2012). As DMP requires data archiving and sharing, data curation and repositories have become another active service area in academic

libraries (MacColl, 2010; Fearon & Association of Research Libraries, 2013; Jones, Pryor, & Whyte, 2013). As of 2013, more than two-thirds of Association of Research Libraries (ARL) libraries provide DMP or related data management services (Tenopir et al., 2014). In addition, the project team conducted a pilot analysis of data services in 31 academic libraries for this project. Our pilot study reveals that a range of services are currently offered in relation to data management, including file organization, data description and storage, data citations, and data management training.

While data management services greatly support researchers with DMP preparation and related data archiving and sharing activities, there has been relatively little effort to understand in what other situations researchers require data services and what services they would most benefit from. Data management might not be a priority for many researchers because it does not directly engage in their main research activities (Markauskaite *et al.*, 2012). Other data services might be also needed that are directly useful for their research projects, such as data collection, analysis, and visualization. This project plans to explore potential areas, in addition to the more typical data management support, where academic libraries might better support researchers' research data needs. This will be accomplished by surveying both researchers (patrons) and service providers (academic libraries).

Significance of the Proposed Planning Project

Arguably, the most important responsibility of libraries in research institutions is to offer research support to researchers. To do this, it is imperative that libraries understand the types of support that researchers need and expect from the library. While an increasing number of libraries offer support in data management, they must determine if that is sufficient, and if not, what other data services they should offer. This project team will survey researchers in order to investigate their general data service needs and to identify when and in what situations they might benefit from expanded services. This will enable libraries to expand the scope of their data services to be more actively engaged in the research process, and the scope and types of data services should be determined from the assessment of patron needs. Additionally, the project team will survey librarians in order to understand their perceptions on RDS in libraries, and to identify types of knowledge and skills needed to provide these services. Additionally, this project will suggest curricula and content needed for formal training of data services librarians.

The findings from this investigation of various stakeholders will be used to develop a conceptual model of RDS. This model will include a list of potential services, required associated resources, skills needed for service providers, and other useful information. It will serve as a reference that libraries can consult in developing, upgrading, and deploying their own services.

IMPACT

Filling the Gap

The project team has identified three gaps in current RDS practice and research: (1) existing services focus on DMP and data curation, rather than direct, embedded research support; (2) there has been limited discussion on the formal education of data services librarians, particularly in the area of curriculum development; (3) there are few reference models that describe the various research data services and associated components. To fill these gaps, the proposed project will address the issues of (1) potential research support based on a need assessment, (2) education of data services librarians aimed at building greater skills in the library workforce, and (3) developing a conceptual data services model that libraries can adopt as a reference model. The project will impact RDS practice across the United States by filling the gaps in existing services. It will suggest a wider variety of services that will meet diverse need of researchers. In addition, this project will initiate more active discussion on the formal education and training of data services librarians in the community of LIS schools.

Project Outcomes

This planning project will establish a baseline assessment of current practice and potential patron needs for RDS. In addition, the project team will solicit opinions from a diverse group of stakeholders, including patrons, librarians, and LIS scholars. They will then create a conceptual model of RDS that will serve as a theoretical, foundational framework for the development of these services in academic libraries. The outcomes expected

from this project will (1) be useful to inform academic libraries as they plan data services that reflect the diverse needs of patrons and (2) offer a guideline of data service-related curricula in LIS education. The project will generate the following final outcomes:

- (a) List of data services currently offered in the US (including select best cases and practices)
- (b) List of situations in which patrons need RDS
- (c) List of feasible RDS that may be offered by academic libraries
- (d) List of resources needed to offer specific RDS
- (e) List of knowledge and skills required for data services librarians
- (f) Strategies to deploy RDS in academic libraries
- (g) Curricula suggestions for data-related LIS programs
- (h) Conceptual model of RDS in academic libraries

Evidence of Project Success

This project is closely related to the **National Digital Platform** movement as it deals with research data in a digital format, and its curation, access, and sharing throughout the United States. It will have an impact across multiple disciplines that use empirical data in a digital format, ranging from the humanities, social sciences, business, natural sciences, health sciences, and engineering. In addition, this project addresses the issue of **Learning in Libraries** as it concerns training and education of librarians to successfully perform RDS. The project will facilitate building and enhancing skills and abilities in the library workforce through relevant training and education in the LIS field.

The project team expects that the outcomes listed above will be adopted by many libraries that currently offer or plan RDS. First, the findings from the case studies and need assessment will be useful for libraries to plan, design, or upgrade new services. Second, the results from the surveys of librarians will inform strategies of operating various data services and managing related resources. Third, results from the interviews of LIS scholars can be adopted by LIS programs that plan to develop new courses related to data services. Fourth, the conceptual model of RDS will function as a reference model to both academic libraries and LIS scholars.

As this proposal is a planning project, it will produce an impact far beyond the immediate success of the project itself. The resulting conceptual model will be used in a subsequent project to develop an RDS prototype, which will be implemented, tested, and enhanced in an operating academic research library. The RDS model and practical guidelines will facilitate data services practice across research universities in the United States and will lead to a significant advancement of nation-wide research support capabilities in the emerging computational science environment.

Evaluation Plan

The project will undergo three forms of evaluation. First, as this project includes research components, the project team will examine the validity and reliability of the data collected from case studies, surveys, and interviews. Inter-coder reliability will be checked for open coding and content analysis in the case studies. For survey and interview questionnaires, an advisory board will check whether questionnaire items are appropriate to achieve the research objectives to ensure the validity of the data. Internal reliability will be checked for survey responses. Second, focus groups will be used to evaluate and ensure the feasibility of data services that are suggested from this project. In addition, the advisory board will be asked to give comments regarding the feasibility of data services. Third, success indicators will include the impact of the project outcomes and research products. The project team will share the project information, preliminary findings and other products on a project website (see the Communication section), and will check the number of visits and downloads of products. The advisory board will also examine the usefulness and potential impact of the project outcomes. Additionally, a long-term indicator of success will be measured from the number of citation and circulation of publications as a way to indicate the success of the project. The project team will also collect feedback and comments in general from those interested in the project via the project website, social media, and conference presentations.

PROJECT DESIGN

To address the research questions proposed, multiple research methods will be employed, such as case studies, surveys, interviews, and focus groups. Table 1 lists specific research questions, associated methods and outcomes.

Table 1. Research Questions

Time	Research Questions	Methods	Outcomes*
Month 1-6	- What kinds of RDS are currently offered in academic libraries? - What resources are used for these services?	Case studies	(a)
Month 1-8	- In what situations do patrons need RDS? - What RDS are patrons most likely to use, least likely to use? - To what extent will patrons use RDS, or would like to use RDS? - What new RDS would patrons like to see in the future?	Survey of patrons	(b) and (c)
	- To what extent are librarians/administrators aware of the importance of RDS? - What RDS are currently offered? What are the roles of the librarians who offer these services? - What relevant skills do librarians currently have to offer RDS? - What new RDS do librarians think possible in the future? - What are the challenges and opportunities in RDS from the librarian perspective? - What are effective ways to deploy and offer these services?	Survey of librarians	(a), (c), (d), (e), and (f)
	- What data-related knowledge and skills do data librarians need? - What subjects and content need to be taught to educate data services librarians?	Interview of LIS scholars	(e) and (g)
Month 8-11	- What kinds of RDS are the most and least useful for patrons? - What services are feasible in current academic libraries? - What resources are needed to offer these services? - What are effective ways to deploy and offer these services?	Focus group	(c), (d), (e), and (f)
Month 9-12	- What specific services should comprise RDS in libraries? - What resources are required for each type of service? - What skills are required to provide these services?	Consolidating findings from the project	(h)

* See Project Outcomes in the Impact section

1. Case Studies (Month 1 – 6)

Schedule: Coding scheme development, and Coder training (Month 1-2), Content analysis (Month 2-5), Case study result (Month 6)

Existing services for research data will be examined through intensive case studies. The purposes of the case studies are: (1) to identify RDS currently offered; (2) to identify resources to be used for RDS; (3) to find best cases and practices for benchmarking.

Different types of academic libraries will be selected for case studies in order to examine the effect of variance caused from size. Carnegie Classification of Institutions of Higher Education will be used to select case study samples (<http://classifications.carnegiefoundation.org/>). As RDS are more widely serviced in large research libraries, more samples will be selected from doctorate-granting R1 and R2 universities. Our initial search showed that only a small number of Master's Colleges and Universities or Baccalaureate Colleges offer RDS. Associates Colleges are excluded from the analysis because RDS is designed for research support in research institutions, rather than teaching colleges. From the list of the institutions, 150 universities or colleges will be randomly selected. To consider diverse cases, we will select several of historically black colleges and universities (HBCUs), colleges in the Appalachian region, women's colleges and universities, and/or Hispanic serving intuitions (see the Diversity Plan section). If the selected institution does not provide any RDS, we will choose another institution randomly until the planned sample number as Table 2 below.

Table 2. Sample selection for case studies

Category	Type	Sample size
Doctorate-granting Universities	Highest Research Activity (R1)	70
	Higher Research Activity (R2)	40
	Moderate Research Activity (R3)	20
Master's Colleges and Universities		10
Baccalaureate Colleges		5
Diversity Cases (see the Diversity Plan section)		5
Total		150

The project team will create a coding scheme based on open coding method, and the consultants will review the coding scheme to ensure its validity. The coding scheme will guide the coders to analyze the websites and extract required data for the case study. The project team already conducted a pilot case study with 31 doctorate-granting R1 universities, and here are some potential services that we are going to investigate in the case study: DMP for external grants; data file management; data storage and sharing; data references services and access; data management training and/or workshops; intellectual property consulting; data analysis; data presentation and visualization; data analysis workshops, tutorials, and/or lectures; and others.

For each service, the following specifics will be examined: (1) whether the service is currently provided or planned; (2) if the service is provided, what is the official name of the service, and what are the content of the service (e.i., specific description of the service?); (3) who are the potential audiences of the service?; (4) which department or team is in charge of the service?; (5) what are the titles of librarians/staff in the department/team?; (6) what resources are associated with the service (e.g., DMP tools?); and others.

Two graduate students from the University of Kentucky LIS program, who are in the data science track, will be hired to analyze the selected websites. The coders will be trained, and conducted initial analysis around 5% of sample cases under the supervision of the PIs to ensure the reliability of the data collection. Each coder will analyze about 20 websites per month, and for any case unclear to code, the PIs and the coders will discuss that case to solve and reach an agreement. We will create a web database server for the coders to work online and share the collected data with the PIs in real-time base. In this way, the PIs will keep track of the quality of case study data. Once all case study data are collected (by Month 5), the PIs and consultants will identify the status of existing RDS. More importantly, approximately top 10% good service cases will be selected by the PIs, and then, the consultants and the advisory board members will review and recommend a final list of best cases of RDS.

2. Survey of Patrons (Month 1 – 8)

Schedule: Survey design and IRB (Month 1-3), Data collection (Month 4-5), Data analysis (Month 6-8)

A survey will be administered to researchers (potential patrons) likely to use RDS from different disciplines. The purposes of this survey are: (1) to identify in which situations researchers need RDS; (2) to identify what kinds of RDS researchers would like to use, and to what extent they would use RDS?; (3) to investigate what new RDS (not currently offered) users are interested in; and (4) to analyze the similarities and dissimilarities of the results by discipline.

The PIs and consultants will draft an initial survey questionnaire. It will be reviewed by the advisory board to assure the validity of survey items. Based on their comments, the survey will be enhanced and pre-tested with at least 10 potential subjects. The survey items will be designed to investigate: in what situations do patrons need RDS?; what RDS would patrons like to use?; to what extent would patrons use RDS?; what new RDS would patrons like to see in the future?; and others. In addition, we will explore various contextual and demographic factors influencing their use intention of RDS, such as awareness, attitude toward library services, data analysis skills, and research domains.

Due to the limited time allowed for the planning project, the University of Kentucky (UK) has been chosen as a representative population who will potentially benefit from RDS. UK is one of the doctorate-granting universities with highest research activity (R1) with about 30,000 students and 99 master programs and

66 doctoral programs. The survey will be distributed to potential patrons of RDS at UK, including faculty, scientists/researchers, and graduate students.

The project team obtained a list of most recent email addresses for the entire faculty, researchers, and students at UK from the Registrar's Office. Using the UK Qualtrics system, a survey will be administered online (<http://www.uky.edu/ukat/atg/qualtrics>). In total, an online survey invitation will be distributed to approximately 12,000 patrons, including about 2,000 faculty, 3,000 scientists/researchers including post-docs, and 7,000 graduate students. A response rate of approximately 7% is anticipated based on a similar survey carried out at UK, while will result in around 840 completed responses. To facilitate survey participation, an incentive in the form of a \$50 Amazon.com e-gift card will be awarded to 10 randomly selected subjects upon completion of the survey.

The collected data will be analyzed both quantitatively and qualitatively as survey items will include likert-scale questions, categorical questions, and open-ended questions. The survey results will reveal in what situations and in which research stage patrons would need support from the library with regard to research data. Taxonomies will be specified and categorized into different types of situations where RDS are needed, and each type of situation will be matched with relevant data services.

In addition, we will analyze the most and least popular potential RDS that patrons indicate a need for. We will list potential types of RDS available in the library, and for each type of RDS, participants' willingness to use that service will be assessed to determine the importance of that service from the perspective of the patrons. The similarities and dissimilarities of patron needs will be examined across different disciplines. In addition, we will look into how patron needs differ among faculty, researchers/scientists, and graduate students.

3. Survey of Librarians (Month 1 – 8)

Schedule: Survey design and IRB (Month 1-3), Data collection (Month 4-5), Data analysis (Month 6-8)

A survey will be administered to librarians who are involved in data services in academic libraries. To ensure the validity, the survey items will be reviewed by the advisory board. The purposes of the librarian survey are: (1) to identify current and potential RDS; (2) to identify resources required for these services; (3) to define the roles of librarians in RDS; (4) to identify knowledge and skills libraries have or need to offer RDS; (5) to investigate the challenges and opportunities in RDS from the library side; (6) to solicit the opinions about effective RDS deployment; and others.

The project team has collected email addresses of about 3,000 academic library deans or directors in the United States. This list will be used to distribute the Qualtrics questionnaire. In the invitation email, we will ask the dean or director to forward the invitation to their librarians who are involved with various RDS. Around 300 completed responses are expected after multiple reminders. To encourage survey participation, an incentive in the form of a \$50 Amazon gift card will be given to 10 randomly selected participants who complete the survey. Similar to the researcher survey, the collected data will be analyzed quantitatively and qualitatively. From the analysis, we will identify which services are most widely offered and what roles librarians perceive important in relation to RDS. Also, the analysis will identify the required knowledge and skills for each type of data service. As this planning project is an exploratory study, qualitative analysis based on open-ended questions will be also important. Moreover, challenges and opportunities will be investigated from the perspectives of librarians based on both quantitative and qualitative analysis to come up with expected implications for RDS. Moreover, the differences of RDS will be examined by different types of institutions (e.g., Carnegie Classification categories, library size, number of staff dedicated to data services, etc.)

4. Interviews with LIS Scholars (Month 1 – 8)

Schedule: Interview question design and IRB (Month 1-3), Data collection (Month 4-5), Data analysis (Month 6-8)

In an effort to suggest educational curricula for data services librarians, we will conduct in-depth interviews with 8 scholars in ALA accredited LIS scholars. The purpose of these interviews is to suggest courses and curricula to educate data services librarians in ALA accredited LIS programs.

The project team will invite eight scholars, particularly faculty in LIS programs who have studied RDS and/or taught data science courses for library science students, to participate. Potential participants will be

selected from those who made relevant journal articles or conference presentations in RDS related topics. Our questions will include: (1) what are the key skills and knowledge required for data services librarians?; (2) what subjects and content should be covered to educate data services librarians?; (3) how many credit hours would be appropriate for data science related courses to certify a data services librarian concentration or track?; (4) what should be prerequisite for data science courses for library science students?; (5) what is the best way to promote the data services librarian concentration or track?; and (6) are there any other factors to consider in educating data services librarians in LIS programs. The interviews will be a form of semi-structured, which relies on predefined questions to collect responses for the identified research questions and the interviewer can further investigate deeply via probing questions. Two interviewers will be hired from the UK LIS program, and they will be trained under the direction of the PIs to assure the quality of interviews. Each interviewer will perform four interviews between Month 4 and 5. All interviews will be conducted via phone, and each interview session will be around one hour. As a token of compensation, a \$50 online gift card will be given for each participant.

The collected data will be analyzed qualitatively using open coding and content analysis. For curriculum development, a list of subjects/topics will be extracted from the analysis of the interview transcripts, and potential courses will be suggested to cover those identified subjects. Also, required knowledge and skills will be mapped with those courses as a guideline for data services librarian education.

5. Focus Groups (Month 8 – 11)

Schedule: IRB (Month 8), Round 1 (Month 9), Round 2 (Month 10)

Once we complete the surveys with patrons and librarians, we will administer two rounds of focus group interviews. The purposes of the focus group interviews are: (1) to prioritize existing and potential RDS (to determine most important, useful services); and (2) to determine the feasibility of services; and (3) to identify resources needed to offer RDS.

The focus group will include a panel of eight members, including two academic librarians, two LIS scholars, and four potential researchers from four different disciplines (e.g., social sciences, humanities, natural sciences, engineering). In order to recruit panel members, we will ask for interested participants in the surveys outlined above. Prior to the initial focus group, each panel member will be asked to review preliminary findings, including a list of existing and potential RDS, from the case studies and surveys of patrons and librarians. In particular, they will be asked to rank the identified services in terms of importance and usefulness before the first meeting. The two rounds of focus group meetings will occur online via UK Adobe Connect (<https://www.uky.edu/ukat/atg/adobeconnect>) in months 9 and 10 respectively. Each meeting will last around two hours, and the project team will preside the meetings. During the first meeting, the focus group will discuss which RDS are important and useful, and then, rank them. That is, we will determine what kinds of research data services are deemed most and least important and useful for patrons. This focus group will also determine the feasibility of each service by reviewing required resources required for that service. The second round focus group will go over the ranking again, and finally decide which services should be included in the conceptual RDS model based on the analysis of importance, usefulness, and feasibility. In this way, the focus group will perform a key role to create a conceptual RDS model in this project.

6. Designing a Conceptual Services Model, Dissemination of the Outcomes, and Planning a Larger Project (Month 9 – 12)

Schedule: Initial draft of a model (Month 9), Enhancement and revision (Month 10), Final version of the conceptual RDS model (Month 11), Final report writing (Month 10-12)

Finally, the PIs and the consultants will create a comprehensive conceptual RDS model by consolidating the findings from case studies, surveys, interviews and focus groups carried out from Months 1-8. As described earlier, the conceptual model will include a list of potential services, associated resources and required librarian skills, and other information. The project team will write a report to submit to IMLS and prepare to disseminate the findings and outcomes of the project through multiple channels. The team will also start working on partnerships with potential partners (e.g., UK Libraries) for a future project. See the Communication and Sustainability sections for a detailed discussion.

DIVERSITY PLAN

Diversity issues will be addressed throughout the project in multiple ways: First, in the planned case studies, we will include several of historically black colleges and universities (HBCUs), colleges in the Appalachian region, women's colleges and universities, and/or Hispanic serving institutions to investigate data services status in the institutions for the underrepresented groups. Second, in the survey of librarians, we will ask in the open-ended questions what kinds of efforts they have made to promote research data services to diverse groups of patrons. Third, in the interviews with LIS scholars, we will ask in what ways diverse groups can be more involved in data services education and practice in librarianship.

PROJECT RESOURCE: PERSONNEL, TIME, BUDGET

Dr. Soohyung Joo, Principal Investigator, is an assistant professor in the School of Information Science at the University of Kentucky. He has been actively involved in the research of data science and services for several years. He made a number of publications in leading LIS journals, such as *Journal of the Association of Information Science and Technology* and *Information Processing and Management*, with the topics of information retrieval, data science, digital libraries, data mining, and online information use. He is actively engaged in developing a data science track in the UK LIS program to prepare future data services librarians. This project is a natural extension of his research agenda to design research data services in academic libraries.

Christie Peters, Co-PI, is Head of the Science & Engineering Library and Coordinator of eScience Initiatives at the University of Kentucky Libraries, as well as a PhD student in the College of Communication (within which resides the School of Information Science). In her role in the UK Libraries, Christie is coordinating efforts to develop and provide research data services within the library system to support the research enterprise at the University of Kentucky. She is also chairing a campus research data advisory group that includes participants from various units and colleges on campus in an effort to develop data-related policies and coordinate research data services across campus. Christie plans to focus on the development and implementation of data services in libraries in her doctoral work.

To ensure the success of the project, our team has invited two consultants and four advisory members. The consultants and advisory members are leading experts in a variety of areas, ranging from information user studies, information need assessment, knowledge management, scholarly communication, data analytics, data repositories, data sharing and reuse in e-science, to data science education.

The responsibilities of the consultants include: reviewing the validity of coding scheme for case studies; helping to design survey items and interview questions; assisting in the creation of an RDS model; and providing feedback on request of the PIs.

Dr. Lisa O'Connor, Consultant, currently an Associate Professor in the UK LIS program with expertise in user behavior and assessment. Her research focuses on user behaviors within virtual and physical environments, particularly in the area of personal financial management and civic engagement. Her previous work, Project SAILS, a standardized tool to assess information literacy skills, was funded by IMLS and adopted by the Association of Research Libraries as a new measures initiative. She will become Chair of the School of Library and Information Studies at the University of North Carolina – Greensboro on August 1, 2016.

Dr. Sean Burns, Consultant, is an assistant professor in the School of Information Science at UK, where he teaches in the library and information science and the information and communication technology programs. His research focuses on the intersection of science communication and knowledge management and he has a special interest in the data and information sources and services that academic librarians provide to researchers.

The advisory board members will provide feedback on research design and outcomes throughout the project and offer general guidance on project activities. In particular, they will be asked to review the survey questionnaires and the outcomes of the project. The advisory members are: **Dr. Kun Lu**, School of Library and Information Science at the University of Oklahoma. Dr. Lu has expertise in data science, data mining, and applied informetrics; **Dr. Youngseek Kim**, School of Information Science at the University of Kentucky. Dr. Kim's research areas are data sharing and reuse, data repositories, and curriculum development for information professionals; **Robert Shapiro**, Assistant Director for the Research, Education & Clinical Services Division, Medical Center Library at UK Libraries; and **Adrian Ho**, Director of Digital Scholarship at UK Libraries. For planning the next larger project that will partner with UK Libraries, two departmental director level librarians

are invited from the UK Libraries as an advisory board member. This planning project has also received support from the School of Information Science (SIS) at UK and the UK Libraries.

The total amount requested from IMLS is \$49,844. Although cost sharing is not required for the Planning Project category, the School of Information Science and the UK Libraries will provide support in several ways, such as management of project finances, online survey tools, web spaces, and institutional repository. Table 3 presents project management plan. Please see the Budgetjustification.pdf and Scheduleofcompletion.pdf for details.

Table 3. Project management plan

Activities	Time	Major objectives	Budget (est.)
Case studies	10/01/16 – 3/31/17	Understand the current status of RDS and define best practices	\$6,108
Patron survey	10/01/16 – 5/31/17	Assess the need of patrons	\$8,556
Librarian survey	10/01/16 – 5/31/17	Understand the status of RDS practice in the field	\$8,556
LIS scholar interview	10/01/16 – 5/31/17	Identify content to be covered in curricula for data services librarian education	\$8,706
Focus group	05/01/17 – 8/31/17	Prioritize existing and potential RDS services	\$5,566
Data services model/ Dissemination	06/01/17 – 9/30/17	Design a conceptual RDS model	\$12,352
Total			\$49,844

COMMUNICATIONS PLAN

Multiple channels will be utilized to communicate with those who are interested in this project and to disseminate the findings of the study. First, a project website will be created to present findings and outcomes, including case study and survey result reports. The website will serve as a communication platform for those who are interested in RDS in the academic library setting. The website will include a blog to share diverse ideas on this topic, as well as to solicit feedback from those interested in the project. A Twitter account will be incorporated into the website to easily reach out individuals and entities interested in the project. Second, the project team will present ideas and findings from this project at professional conferences, such as ALA and Research Data Access & Preservation Summit (RDAP), as well as at academic conferences, such as the Annual Meeting of ASIST and JCDL. Third, findings of the study will be submitted to relevant journals, such as Journal of Academic Librarianship, Library and Information Science Research, and D-Lib Magazine. Fourth, preliminary findings will be stored and shared in the institutional repository of the University of Kentucky, UKnowledge (<http://uknowledge.uky.edu/>) in pre-print format.

SUSTAINABILITY

This planning project will be extended to two larger projects. First, the findings of this project will be used to develop and test comprehensive data services at the UK Libraries. The project team will initiate a partnership with the UK Libraries and discuss what kinds of RDS can be offered based on the analysis of available resources. The results of the patron needs assessment in the planning project will be used to plan user-centered data services. The Co-PI, Christie Peters, is in a position where she can plan and manage research data services at the UK Libraries, and will coordinate the project to implement and test new research data services. In this way, the conceptual RDS model will be implemented into tangible services. In addition, the project team will continue conducting research on this topic to advance the RDS model and to keep a track of changing patron needs. Second, in order to develop curricula for data services librarianship, the project team will apply for the full Laura Bush 21st Century Librarian Program Project Grant for Master's Programs in collaboration with the School of Information Science at UK. The UK LIS program has offered a series of data science related courses, such as Introduction to Data Science and Data Analysis and Visualization. The needs of RDS professionals are rapidly evolving. The full project will help to educate the next generation of RDS librarians.

DIGITAL STEWARDSHIP SUPPLEMENTARY INFORMATION FORM

Introduction

The Institute of Museum and Library Services (IMLS) is committed to expanding public access to federally funded research, data, software, and other digital products. The assets you create with IMLS funding require careful stewardship to protect and enhance their value, and they should be freely and readily available for use and re-use by libraries, archives, museums, and the public. However, applying these principles to the development and management of digital products is not always straightforward. Because technology is dynamic and because we do not want to inhibit innovation, we do not want to prescribe set standards and best practices that could become quickly outdated. Instead, we ask that you answer a series of questions that address specific aspects of creating and managing digital assets. Your answers will be used by IMLS staff and by expert peer reviewers to evaluate your application, and they will be important in determining whether your project will be funded.

Instructions

If you propose to create any type of digital product as part of your project, complete this form. We define digital products very broadly. If you are developing anything through the use of information technology (e.g., digital collections, web resources, metadata, software, or data), you should complete this form.

Please indicate which of the following digital products you will create or collect during your project
(Check all that apply):

	Every proposal creating a digital product should complete ...	Part I
	If your project will create or collect ...	Then you should complete ...
<input type="checkbox"/>	Digital content	Part II
<input type="checkbox"/>	Software (systems, tools, apps, etc.)	Part III
<input type="checkbox"/>	Dataset	Part IV

PART I.

A. Intellectual Property Rights and Permissions

We expect applicants to make federally funded work products widely available and usable through strategies such as publishing in open-access journals, depositing works in institutional or discipline-based repositories, and using non-restrictive licenses such as a Creative Commons license.

A.1 What will be the intellectual property status of the content, software, or datasets you intend to create? Who will hold the copyright? Will you assign a Creative Commons license (<http://us.creativecommons.org>) to the content? If so, which license will it be? If it is software, what open source license will you use (e.g., BSD, GNU, MIT)? Explain and justify your licensing selections.

A.2 What ownership rights will your organization assert over the new digital content, software, or datasets and what conditions will you impose on access and use? Explain any terms of access and conditions of use, why they are justifiable, and how you will notify potential users about relevant terms or conditions.

A.3 Will you create any content or products which may involve privacy concerns, require obtaining permissions or rights, or raise any cultural sensitivities? If so, please describe the issues and how you plan to address them.

Part II: Projects Creating or Collecting Digital Content

A. Creating New Digital Content

A.1 Describe the digital content you will create and/or collect, the quantities of each type, and format you will use.

A.2 List the equipment, software, and supplies that you will use to create the content or the name of the service provider who will perform the work.

A.3 List all the digital file formats (e.g., XML, TIFF, MPEG) you plan to create, along with the relevant information on the appropriate quality standards (e.g., resolution, sampling rate, or pixel dimensions).

B. Digital Workflow and Asset Maintenance/Preservation

B.1 Describe your quality control plan (i.e., how you will monitor and evaluate your workflow and products).

B.2 Describe your plan for preserving and maintaining digital assets during and after the award period of performance (e.g., storage systems, shared repositories, technical documentation, migration planning, commitment of organizational funding for these purposes). Please note: You may charge the Federal award before closeout for the costs of publication or sharing of research results if the costs are not incurred during the period of performance of the Federal award. (See 2 CFR 200.461).

C. Metadata

C.1 Describe how you will produce metadata (e.g., technical, descriptive, administrative, or preservation). Specify which standards you will use for the metadata structure (e.g., MARC, Dublin Core, Encoded Archival Description, PBCore, or PREMIS) and metadata content (e.g., thesauri).

C.2 Explain your strategy for preserving and maintaining metadata created and/or collected during and after the award period of performance.

C.3 Explain what metadata sharing and/or other strategies you will use to facilitate widespread discovery and use of digital content created during your project (e.g., an API (Application Programming Interface), contributions to the Digital Public Library of America (DPLA) or other digital platform, or other support to allow batch queries and retrieval of metadata).

D. Access and Use

D.1 Describe how you will make the digital content available to the public. Include details such as the delivery strategy (e.g., openly available online, available to specified audiences) and underlying hardware/software platforms and infrastructure (e.g., specific digital repository software or leased services, accessibility via standard web browsers, requirements for special software tools in order to use the content).

D.2 Provide the name and URL(s) (Uniform Resource Locator) for any examples of previous digital collections or content your organization has created.

Part III. Projects Creating Software (systems, tools, apps, etc.)

A. General Information

A.1 Describe the software you intend to create, including a summary of the major functions it will perform and the intended primary audience(s) this software will serve.

A.2 List other existing software that wholly or partially perform the same functions, and explain how the tool or system you will create is different.

B. Technical Information

B.1 List the programming languages, platforms, software, or other applications you will use to create your software (systems, tools, apps, etc.) and explain why you chose them.

B.2 Describe how the intended software will extend or interoperate with other existing software.

B.3 Describe any underlying additional software or system dependencies necessary to run the new software you will create.

B.4 Describe the processes you will use for development documentation and for maintaining and updating technical documentation for users of the software.

B.5 Provide the name and URL(s) for examples of any previous software tools or systems your organization has created.

C. Access and Use

C.1 We expect applicants seeking federal funds for software to develop and release these products under an open-source license to maximize access and promote reuse. What ownership rights will your organization assert over the software created, and what conditions will you impose on the access and use of this product? Identify and explain the license under which you will release source code for the software you develop (e.g., BSD, GNU, or MIT software licenses). Explain any prohibitive terms or conditions of use or access, explain why these terms or conditions are justifiable, and explain how you will notify potential users of the software or system.

C.2 Describe how you will make the software and source code available to the public and/or its intended users.

C.3 Identify where you will be publicly depositing source code for the software developed:

Name of publicly accessible source code repository:

URL:

Part IV. Projects Creating a Dataset

1. Summarize the intended purpose of this data, the type of data to be collected or generated, the method for collection or generation, the approximate dates or frequency when the data will be generated or collected, and the intended use of the data collected.

2. Does the proposed data collection or research activity require approval by any internal review panel or institutional review board (IRB)? If so, has the proposed research activity been approved? If not, what is your plan for securing approval?

3. Will you collect any personally identifiable information (PII), confidential information (e.g., trade secrets), or proprietary information? If so, detail the specific steps you will take to protect such information while you prepare the data files for public release (e.g., data anonymization, data suppression PII, or synthetic data).

4. If you will collect additional documentation such as consent agreements along with the data, describe plans for preserving the documentation and ensuring that its relationship to the collected data is maintained.

5. What will you use to collect or generate the data? Provide details about any technical requirements or dependencies that would be necessary for understanding, retrieving, displaying, or processing the dataset(s).

6. What documentation (e.g., data documentation, codebooks, etc.) will you capture or create along with the dataset(s)? Where will the documentation be stored, and in what format(s)? How will you permanently associate and manage the documentation with the dataset(s) it describes?

7. What is the plan for archiving, managing, and disseminating data after the completion of the award-funded project?

8. Identify where you will be publicly depositing dataset(s):

Name of repository:
URL:

9. When and how frequently will you review this data management plan? How will the implementation be monitored?

Planning Research Data Services in Academic Libraries: Designing a Conceptual Services Model based on Patron Needs Assessment

Project Overview and Need

This planning project will serve as an initial investigation for a larger project aimed at developing research data services in academic libraries to support researchers in data collection, management, analysis, and presentation. The goal of this project is to design a research data services model at the conceptual level based on a needs assessment of researchers in academic libraries, while the larger project is expected to yield a physical level data services prototype as well as specific guidelines for librarian training and LIS curricula for research data services. In this project, *research data services* refers to a range of library services to assist academic researchers to collect, manage, analyze, present, visualize, and distribute data in their research activities. For example, research data services can potentially include a wide variety of services, such as: (a) data collection; (b) data reference services; (c) data refinement; (d) data storage and management; (e) data analysis; (f) big data analysis; (g) data visualization; (h) relevant workshops or tutorials; and other services. The planning project proposed herein focuses on the perceived data-related needs of diverse patrons and stakeholders, such as library patrons, librarians, administrators, and scholars.

In the research community, the paradigm has recently shifted to computational science across different disciplines, which deals with complicated, large datasets that require advanced techniques to manage and analyze them. As computational methods emerge as central to the research community, researchers can expect to benefit from new types of data services offered by academic research libraries, but a better understanding is needed of how these nascent services support rapidly evolving research methods and how to tailor them to various patron needs. This planning project will investigate: (1) current status of research data services; (2) researcher community needs assessment; and (3) perceptions and opinions of heterogeneous stakeholders. Based on the findings from this initial investigation, we will suggest a research data service model at the conceptual level. This *conceptual data service model* will identify types of data services, associated resources necessary for services, service platforms, knowledge and skills needed by librarians, and corresponding librarian training and education plans. This conceptual model will be used in a subsequent project to develop a research data services prototype, which will be implemented, tested, and enhanced in an operating academic research library. The research data services model and practical guidelines will facilitate data services practice across research universities in the United States, and it will lead to a significant advancement of nation-wide research support capabilities in the emerging computational science environment.

The research team consists of: Soohyung Joo (PI), an assistant professor in the UK LIS program, has been actively involved in the research of data science and services for several years. Joo is currently developing a data science track in the UK LIS program to prepare future data services librarians. Christie Peters (Co-PI), Head of the Science Library & eScience Initiatives at the UK Libraries, is leading efforts to develop data services for faculty and students across the sciences and engineering and is collaborating with others in the Library and university system to develop comprehensive data services on campus for all UK students and faculty. Lisa O'Connor (Consultant) is an associate professor in the UK LIS program with expertise in user behavior and assessment. Her previous work, Project SAILS, was funded by IMLS. We expect to find and recruit additional experts for future collaboration on this project.

Project Design, Task Goals, Outcomes and Budget

Table 1 presents the project design, which specifies the research questions, objectives, methods, and outcomes. Multiple methods will be employed to achieve the research objectives, including case studies, surveys, and focus groups.

First, existing services for research data will be examined through intensive case studies. The research team plans to visit approximately 200 top research university library websites to analyze the types of research data services that are currently provided and the resources that are used to provide those services. Second, surveys and interviews will be administered to heterogeneous groups, including (a) researchers from different disciplines, including faculty, scientists, post-docs, and graduate students; (b) librarians engaged in data services (e.g., data services librarians, e-science librarians); (c) library administrators (e.g., Deans, directors); and (d) library science scholars (See the Research Questions for details). For each group, we will investigate their perceptions and opinions, and solicit suggestions for research data services in academic libraries. For example, the researchers group will be asked: in what kinds of situations are research data services needed?; what kinds of research data services do you think would be useful for libraries to offer?; what kinds of research data services have you used before, if any?; *etc.* Third, we will conduct two rounds of focus group interviews with a panel of eight members, including two librarians, two scholars, and four potential patrons. The focus groups will assist the research team to design a conceptual data services model by analyzing the findings from the case studies, surveys, and interviews. Fourth, initial guidelines and content for librarian training and LIS curricula will be suggested based on the surveys and

interviews from librarians and LIS scholars. Multiple channels will be used to disseminate the findings, such as conferences, workshops, and journal publications. The research team will create a website and social media accounts to share the progress and findings of the project.

Table 1. Project Design

Research Questions	Research Objectives	Methods	Outcomes/Deliverables
What kind of research data services are currently offered in academic libraries? What are the resources used for those services? <i>etc.</i>	To identify the status of current research data services and find successful cases and practices.	Case studies (Month 1 through 6)	A list of data services currently offered in the US. Best cases and practices for benchmarking.
In what situations do patrons need research data services? What research data services would patrons like to use?; To what extent would patrons use research data services? What new research data services (not currently offered) would patrons like to see in the future? <i>etc.</i>	To identify in which situations patrons need research data services and analyze the similarities and dissimilarities of patron needs by discipline.	Surveys and interviews with patrons. (Month 1 through 8)	A list of situations when patrons need research data services. A list of types of data services patrons want. Types of data services for different disciplines.
To what extent are librarians/administrators aware of the importance of data services? What kinds of research data services are currently offered? What are the roles of librarians who offer those services? What kinds of relevant skills do librarians have to offer research data services? What kinds of new research data services do librarians/administrators think possible in the future? What are the challenges and opportunities in research data services from the librarian/administrator perspectives? <i>etc.</i>	To identify current and potential research data services, define the roles of librarians in research data services, solicit the ideas of new services related to data and analysis, understand the motivations of librarians and administrators, and identify the challenges and opportunities in data services.	Surveys and interviews with librarians and library administrators (Month 1 through 8)	A list of current and potential research data services from the library point of view. Strategies to deploy research data services in academic libraries.
What kind of research data services would be useful for researchers? What resources are required for those services? What are effective ways to deploy and offer those services? <i>etc.</i>	To identify research data services useful for patrons, as well as to identify resources needed for research data services and the strategies to deploy/offer those services.	Surveys and interviews with LIS scholars. (Month 1 through 8)	A list of research data services from the scholars. A list of resources needed for services. Strategies to deploy research data services.
What data-related knowledge and skills do data librarians need? What subjects and content need to be taught to educate data services librarians? <i>etc.</i>	To identify skills needed for data services in libraries and identify relevant curricula and courses needed to train data services librarians.	Surveys and interviews with librarians and LIS scholars. (Month 1 through 8)	An initial list of knowledge and skills required for data services librarians. Guidelines for data service librarian training. Curriculum suggestion for LIS programs.
What kind of research data services are the most and least useful for patrons? What services are feasible in current academic libraries? What resources are needed to offer those services? What is the effective, efficient way to offer those services? <i>etc.</i>	To prioritize existing and potential research data services, determine the feasibility of services, identify resources needed to offer data services, and suggest effective and efficient methods to deploy and offer the research data services.	Focus groups (Month 8 through 12)	A conceptual model of research data services.

The total amount requested from IMLS is \$47,772: (a) \$19,538 is allocated to the personnel, including salary support for PI and Co-PI, hourly payment for graduate student assistants, and fringe benefits; (b) \$8,000 is requested for a consultant and focus group panels; (c) \$2,800 is assigned for travels for conferences; (d) \$2,700 is requested for research subject incentives for surveys and interviews; and \$16,684 is requested for indirect costs, which are calculated based on the University of Kentucky's federally negotiated rates.