Report from the Research Data Services Strategic Action Committee

June 2016-January 2017

## Introduction

Many academic libraries are working to provide data management services to their researchers and students, especially those in STEM fields (Tenopir 2014). This trend is largely in response to requirements from many funding agencies requiring Data Management Plans as well as greater accountability regarding the management and sharing of research data. While the attitudes and practices about data sharing and data management vary across disciplines (Tenopir et. al 2015), many researchers are sharing their data more frequently.

The Research Data Services Strategic Action Committee was created to develop a comprehensive and cohesive plan for providing research data services through the University Libraries (hereafter “the library”). In addition, the committee was charged to identify key partners to collaborate with, develop a marketing strategy and a plan for assessment. In this report, we describe our approach, report on our benchmarking analysis, and describe the services we identified as important and feasible to deliver through the library. We then describe two proposed organizational structures along with a suggested hiring plan and other required resources. We also outline a marketing and assessment plan.

## Approach

We took an Agile-like approach to developing the services model. While Agile methodologies are usually associated with software development, they are applicable to a wide range of use cases. The Agile approach is useful for several reasons. First, it is focused on customer or user engagement, with the goal of developing products that fit the user’s needs. Second, it emphasizes communication and teamwork. Finally, it promotes adaptability to change by working with small, distinct chunks of work and working on a short feedback cycle. For our purposes, since we were focused on developing a strategy, our work focused primarily on the users and communication while building a framework for future development of the services model.

This Agile approach will also be used in the implementation. Agile methods are designed to break plans up into small, well defined chunks and facilitate communication among the team and promote communication with the intended users. The goal is to allow the team to adapt to changes in both the environment and with user needs. The report that follows should be used as an initial plan. As needs change, or problems arise, changes will be required to adapt.

We also utilized two tools often used in Agile development. First was Github (https://github.com). While primarily a software development and versioning tool, it has several services that worked well for our purposes. It can serve as a sharing platform for many of our project documents with very robust versioning. It also provides a lightweight but robust issue tracking system that allowed us to generate potential services as individual units. It also has an integrated, and well versioned Wiki where we could write many of our working documents online. The second tool we used was Waffle.io (<https://waffle.io/>). Waffle integrates with the issue tracking system of Github and provides a web-based board of cards, also known as a Kanban Board. This allowed us, and future participants to easily view planned and current work, sort them and track their progress.

We gathered and documented user requirements by creating Personas. Personas are fictionalized descriptions of users and other stakeholders for a product. They are often used in Agile development to keep development user focused. Our personas are stored on GitHub wikis (https://github.com/psu-libraries/personas). We used a dedicated repository separate from our other work so that the Personas could be more easily reused in other projects. We created the personas by interviewing researchers from a variety of disciplines and by including previous knowledge from interactions with researchers in a variety of situations. Each persona is typically a composite of one or more people and is written in such a way that services can be created to fulfill user needs.

We also benchmarked research data services at peer institutions. Benchmarking can be used to learn both from the successes and failures of peer institutions. The results of this benchmarking are described below.

Typically, in Agile development, potential system features are written as user stories with the structure of “As a user I would like to have some feature so that I can realize some benefit”. While user stories are very useful, we found them to be too brief, and focused on too small of services for our purposes. We adopted a slightly different approach called Problem Scenarios. In this process, we identify problems that the users encounter. We then propose a variety of potential solutions that would help solve that problem for the user. This has the advantage in that our problems can be more fully described. Also, in many cases, there are multiple potential solutions to a problem and they can be described and discussed separately.

We documented both our user problems and proposed solutions as GitHub issues in a dedicated repository (<https://github.com/psu-libraries/library_data_services_model>). We also documented potential partners as issues. By applying the appropriate labels, the issues are sorted on a Waffle board (<https://waffle.io/psu-libraries/library_data_services_model>). The goal was to provide a mechanism to easily start the work of developing research data services at the library, monitor progress and, to add, remove or reprioritize issues and work as needed.

## Benchmarking

We looked at the data services offered by the following peer institution university libraries focusing on what services are offered, what resources are dedicated to offering these services, and how are those resources organized:

* University of Iowa Libraries
* University of Michigan Libraries
* University of Minnesota Libraries
* Duke University Libraries
* University of Illinois Champaign-Urbana Libraries
* University of New Mexico
* Purdue University Libraries
* Cornell University Libraries

The University of Iowa Libraries does not have a clearly defined organizational model for its data services. Two librarians from the Engineering Library and the library Department of Research and Instruction provide consultative service for [research data management](http://guides.lib.uiowa.edu/data) [(link to services is embedded in text)](http://guides.lib.uiowa.edu/data). Data discovery services are covered by library subject specialists. Analytical services for GIS are provide for by a specialist in the Map Library.

Iowa’s [Digital Scholarship & Publishing Studio](https://www.lib.uiowa.edu/studio/) is a partnership model for developing services for content analysis and digital humanities. However, it is limited since it only it only provides services for affiliated faculty.

University of Michigan’s “[Research Unit](https://www.lib.umich.edu/research-unit)” is broken up into disciplinary teams one of which specifically focuses on data services. These teams report directly to the Associate University Librarian for Research. Michigan’s [research data services](https://www.lib.umich.edu/research-data-services) focus on the following:

* Data Management Planning
* Discovery & Access
* Data Organization & Management
* Metadata & Documentation
* Data Sharing & Publication
* Preservation
* Data Visualization

Data services within the Duke University Libraries is for the most part centralized organizationally in the [Department of Data and Visualization Services](https://library.duke.edu/data/data-visualization) (DVS). This sits organizationally under the auspices of the university (i.e. outside of the library), Dean for Information Technology Services.

There are seven full-time consultants and librarians that report directly to the head of DVS. This includes: two Senior Research Data Management Consultants one for science and the other for the social sciences, a GIS & Map Librarian as well as a GIS assistant, Data Analytics and Management Librarian focusing on issues related to data wrangling and cleaning, and a Data Visualization Coordinator as well as a Data Visualization Analyst. The department also has a graduate student internship and CLIR fellow for data curation in the social sciences.

Service focus include: data sources, data cleaning, GIS/Digital mapping, data management, data visualization and data analysis. Duke Libraries along with North Carolina State University have been on the leading edge of designing physical space for data and visualization services.

Duke libraries also provide services for data driven projects within the humanities through their [Digital Scholarship Services](https://sites.duke.edu/digital/). Organizationally, this reports to a separate Dean for User Services and Collections. The service has a librarian as well as a digital humanities technology consultant. The services include: plan and implement a digital research project, discover and apply digital methods and tools, training for digital approaches in academic work, referral to expertise and resources at Duke.

The University of Minnesota Libraries data services is organized around a collaborative institutional wide and [referral model](https://www.lib.umn.edu/datamanagement/during). There appear to be several different streams of service support for data within the University Library:

[Research Support Services](https://www.lib.umn.edu/researchsupport) where data is considered part of the broader context of research as well as services for a broad population from advanced researcher to undergraduates. The University Libraries do not provide support for statistical analysis, but refer this to higher level statistical support services within the University.

The [Data Management and Curation Staff](https://www.lib.umn.edu/datamanagement) is comprised of a Director of the institutional repository and five data curation specialists: data management specialist a faculty member from the College of Liberal Arts, data storage specialist from the Office of Information Technology, GIS curator from USpatial, and a liaison librarian from Public Health.

Services for Geospatial data are provided by the [Map Library](https://www.lib.umn.edu/borchert) which is staffed with a Head of the library, a soft-money project-based Geospatial Project Metadata Coordinator position, and soft-money half-time Spatial Data Analyst & Curator that also reports to the central geospatial support unit (USpatial), and a library assistant. More complex Geospatial services for the University are coordinated by [USpatial](https://uspatial.umn.edu/) a part of Research Computing, a center in the Office of the Vice President for Research (OVPR).

DASH (Digital Arts and Humanities) is a cross disciplinary collaboration with the University Libraries [Digital Library Services](https://www.lib.umn.edu/digital) and the [Liberal Arts Technologies and Innovation Services](http://latis.umn.edu/). This allows researchers to explore: Text and data mining, Digital archives, GIS and spatial research, Data visualization and data arts, Digital storytelling, Mobile app development, Multimodal scholarship, Desktop Fabrication, 3D Printing, and Makerspaces, Critical code and algorithm studies

Cornell University Libraries participates in several ways in Cornell’s [Research Management Services Group (RMSG)](http://data.research.cornell.edu/) which is a collaborative institutional-wide model with a formalized structure: consultants, management council, and faculty advisory board.

The RMSG is comprised of 13 consultants outside and within the University Libraries. Seven of those consultants are from the libraries: a life science librarian, medical librarian, environmental science librarian, GIS specialist, project manager for Cornell’s University Library Archival Repository, Digital Projects Librarian, and a Data Curation Specialist.

The coordinator from the RMSG that meets with a Management Council is also from the University Libraries. The council is comprised of stakeholders from Information Technology, Center for Advanced Computing, Astronomy, Institute for Social and Economic Research, Medical College, the Director for Information Technologies within the University Libraries, and the Assistant Unit Leader Scholarly Resources and Preservation Services.

The faculty advisory board is comprised of nine departmental faculty primarily from natural science disciplines, and two from the social sciences. There are ex officio members from the Office of Sponsored Programs and the Office of Research Integrity and Assurance.

The RMSG service areas include:

* Collaboration tools
* Data collection and analysis
* Data sharing
* High performance computing
* Intellectual property and copyright
* Metadata
* Security, privacy, and confidentiality
* Storage, backup, and recovery

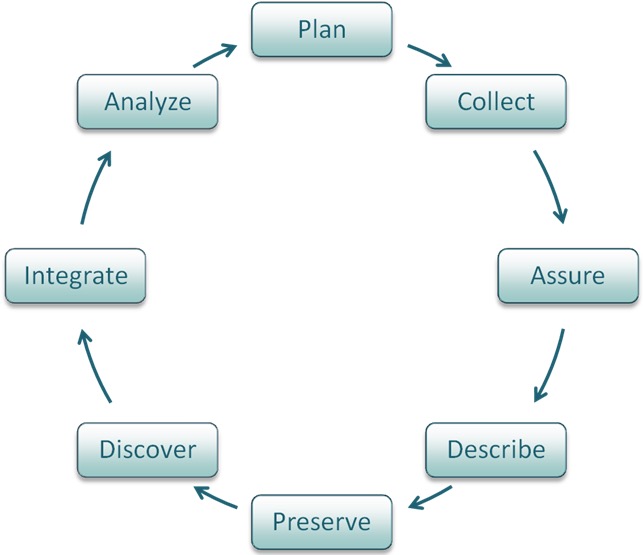


Fig. 1 DataONE research data life cycle. This life cycle was developed to assist DataONE in developing and organizing data services.

## Services Catalog

The services catalog presented here is organized around the data life cycle based on the one proposed by DataONE shown above. Our discussion deviates somewhat from the above life cycle in the interest of brevity as well. This is useful as most activities around data occur primarily during one segment of the lifecycle. However, it should be noted that the services discussed here may at times be used at any point during the data lifecycle. Additionally, the services described here should not be considered complete. In keeping with an Agile process, the data services provided should be continuously evaluated for their effectiveness and modified or even removed as needed. Additional services will also be added as needed.

### Plan

Since 2011 NSF has required the submission of a two-page data management plans (DMPS) along with research proposals. The purpose of this requirement is to 1) ensure that data is properly managed and maintained; 2) promote the sharing of data; 3) ensure that data is available after the conclusion of the project. Additionally, many researchers find that they spend far too much time managing their data, often feel overwhelmed by the prospect of data management and often fail to properly maintain and document their data. The following services are designed to help researchers overcome these difficulties and better fulfill the requirements of their funders.

Data management plan support is a crucial service already offered here and by many other peer institutions. An initial level of support for writing DMPs is the DMPTool offered by the California Digital Library. The DMPTool provides templates for most if not all funding agencies that required data management plans. The templates are augmented with guidance about the various sections in the data management plan for that agency. The DMPTool can also be customized by the partner institutions to provide researchers with institutional specific guidance as well. To fully realize the benefits of the DMPTool it will be integrated with our LibGuides where we will update the pages to provide additional guidance for the various DMP sections.

In addition to the DMPTool, we will provide consultation as needed, providing guidance and additional suggestions to researcher’s DMPs. We are already offering this service and with additional marketing we expect it will be a popular service.

While writing data management plans is important, one of the goals of writing such plans is to help researchers plan early for maintaining, documentation and ultimately archiving their data. We will update the current LibGuide data management pages with useful guidelines and best practices for data management. Topics may include but are not limited to naming conventions, file organization, versioning and data modeling. In addition, we will link to other resources such as the DataONE website that hosts best practices for researchers.

We will also provide workshops for researchers on more advanced topics of data management. University of Minnesota holds a two-day data management boot camp for graduate students. University of New Mexico held a three-week intensive data sciences course for graduate students in the summer. The boot camp exposed students to the basic concepts of data management and gave them exposure to the various resources available to researchers. The three-week course also gave students skills in topics such as website design, relational databases, reproducible science and geospatial analysis. We will also offer a variety of workshops that give researchers exposure to more advanced techniques in data management. For instance, GitHub workshops have been very popular at other Institutions. These workshops are also an excellent opportunity to partner with campus partners such as IST and DataCommons.

While data management planning is often thought of as writing a Data Management Plan, it is better to approach it as preparing to address all aspects of the data life cycle. Therefore, many of the services listed below will also be addressed in data management planning.

### Collect and Integrate

One of the primary reasons for the focus on data management and archiving is the reuse of data. Thus, the task of finding data has become more common. Data reuse varies widely among disciplines. The social sciences, for instance, frequently use census data or survey data stored in a variety of repositories. Genomics data are also commonly shared through the NCBI Genbank and other services. The reuse of data brings new difficulties for which the libraries can provide services.

First, the finding of data. Data repositories can be very domain specific so that librarians who are familiar with the typical data needs for their domain are needed. The libraries can be of considerable value in this realm. Additionally, some data are offered commercially, we must therefore build data into our collections much as we do with print material. Reusing data also brings with it the need of understanding Intellectual Property so that people who use existing data are doing so without running into legal or ethical issues.

### Quality Assurance

Quality assurance can have multiple meanings. For many researchers, quality assurance (QA) involves ensuring that data values are accurate and without error. From the libraries perspective, the meaning also includes ensuring that the data and metadata conform to various standards both external and internal. For instance, date values can have many formats, however researchers often use ad-hoc formats or use the default formats used by their software. The resulting values can often be difficult to interpret unambiguously. Additionally, file formats might not be the best for long term preservation.

The library data services should provide guidance to help researchers create the data that is of high quality both in terms of the values of the data, but also in terms of the structure and format of the data. The workshops mentioned previously would work well to convey this information. Developing online material on our LibGuides and linking to other online resources would work well.

### Describe

Data documentation and metadata are considered one of the most important aspects of data management. Here we are using the term **Documentation** to refer to non-machine readable information, **Metadata** refers to machine readable information, typically standardized. Both terms refer to information that describes other content. The OAIS specification (ISO 14721) describes several types of description. Table 1 shows some of the more commonly discussed types of information in the OAIS specification

|  |  |  |
| --- | --- | --- |
| Metadata/Documentation Type | Description | Example |
| Descriptive Information |  |  |
| Reference Information | Identifies the Content Information. Includes identifiers that allow outside systems to refer unambiguously to a Content Information | DOI  ISBN  ORCID |
| Provenance Information | The information that documents the history of the Content Information. | PROV  PREMIS |
| Context Information | The information that documents the relationships of the Content Information to its environment. This includes why the Content Information was created and how it relates to other Content Information objects. |  |
| Fixity Information | The information which documents the mechanisms that ensure that the Content Information object has not been altered in an undocumented manner. | MD5  SHA1 |
| Access Rights Information | The information that identifies the access restrictions pertaining to the Content Information, including the legal framework, licensing terms, and access control. | Creative Commons  Apache 2.0 |
| Representation Information | The information that maps a Data Object into more meaningful concepts. |  |
| Semantic Information | The Representation Information that further describes the meaning beyond that provided by the Structure Information. |  |
| Structural Information | The Representation Information that imparts meaning about how other information is organized. | File Structures  METS  FITS documentation |

Table 1. Common types of metadata

The purpose of the OAIS specification is to ensure that content is discoverable, findable and that both the content and the meaning of data are preserved. A scan of most data repositories reveals that most data receives adequate fixity and reference information. Additionally, there is typically sufficient descriptive information to make the data discoverable. Most repositories provide services to aid with access rights information, however users are often uncertain about how to use these services. However, most content is not described sufficiently in terms of context, representation, provenance and structure. These categories typically provide information that allow subsequent users to understand and reuse the data.

Because this information is so tightly associated with the specific data set, typically it is the data creators who hold this information and it is considered a best practice for the creators of data to document their data as it is created. However, they often lack the skills, motivation, and resources to sufficiently translate this information to the repository. Curation at the time of deposit into a repository is discussed in a later section. However, it should be noted here that curating even modest sized data sets at the end of the project can be time consuming and data curators will typically lack knowledge specific to a project. Domain knowledge and curation skills are important to ensure that datasets are curated properly and to facilitate communication with the creators.

Data services provided by the library should aim to help researchers document their own data as it is created, and create specific metadata where required. The library data services team will provide guidance to researchers, ideally prior to the start of research to help data creators begin documenting their work from the beginning. We will provide additional support as needed throughout the research cycle, and provide additional support as the data is archived to ensure that it is documented as completely as possible. Specific services would include, for example:

1. Help with writing effective README files.
2. Help with metadata standards. When needed, we will provide support for commonly used standards. Additionally, we will provide guidance for tools that help researchers write metadata files when available. This will also be an excellent opportunity to partner with liaison librarians.
3. Guidance with Licensing and Intellectual property.

### Preservation/Sharing

The primary goal of most of the previously discussed services is to end with research products that are of high quality and well organized with the goal of preserving and sharing the data with future researchers. The libraries in most universities have taken the lead in developing this data service.

The Libraries already offers repository services with ScholarSphere that is suitable for smaller data sets. The ScholarSphere team has also been working to increase its ability to handle larger files and more file rich data sets. Larger data sets are still problematic and present an opportunity to collaborate with other partners such as the Institute for Cyberscience Advanced Cyberinfrastructure to develop user friendly methods for ingesting and hosting large complex datasets. In addition, there are other repositories on campus that the libraries can work with such as Data Commons. The libraries are already working with Data Commons to increase their level of preservation associated with their content. The collaboration should continue and expand when needed to other campus resources to more fully serve the research community at Penn State.

In addition to repositories, the library should take an active role in helping to curate research data. A cursory search through most data repositories reveals that much of the content is poorly documented and organized, drastically reducing the usefulness of the content. By providing curation services via consultation with researchers starting at the beginning of the research, the quality of the resulting data products will be increased.

In addition to Penn State repositories, many other disciplinary repositories exist, such as GenBank, and ICPSR for example. The libraries can assist researchers in identifying appropriate discipline-specific repositories. In addition, many discipline-specific repositories only handle certain types of data. For instance, GenBank can handle sequence data, but will not include other data such a field data, images or other types of data that might have been generated as part of the research. The libraries can work with researchers to ensure that the appropriate data is sent to discipline-specific repositories, but that the additional data is also deposited in an appropriate location.

For many researchers, archiving and sharing data is difficult due to the potential of exposing personal or private information or releasing other forms of sensitive data. The libraries can assist researchers by consulting with them about their data, reviewing their data before it is ingested into the repository, and providing online guidance about managing personal and sensitive information.

### Analysis

The libraries currently offer statistical and visualization consulting through the Data Learning Center. While primarily focused on undergraduate students, graduate students and some faculty have also been served. Analysis and visualization is of course an integral part of the research process and should be continued. In addition, we can offer more advanced topics, workshops perhaps in collaboration with other units on campus.

The libraries currently offer maps (print and digital) and geospatial support services through a new maps and geospatial information unit (beginning January 2017). Geospatial services were formerly associated with the Data Learning Center where they served patrons from every academic college on campus with Liberal Arts, Agricultural Sciences and Earth and Mineral Sciences being the most common home departments of patrons. Requests for geospatial services have grown rapidly, increasing by 50% - 100% each semester since the Fall of 2014. Patrons are primarily graduate students, but also faculty, undergraduates and postdocs in that order of frequency. Maps and geospatial services include one-on-one reference and research consultations, for-credit course integrations in partnership with teaching faculty, training of undergraduate interns and graduate research assistants, participation in consortial projects, workshops, outreach events, maps and geospatial data collection development and maintenance, and geospatial data creation.

## Organization and Resources

Data management service programs vary considerably into their structure. Many libraries cobble together a program by adding additional duties to existing library positions, like. liaison librarians. Other libraries take a more formal, centralized approach, devoting additional resources with some staff being entirely focused on providing data management services. Many also have dedicated research data librarians that focus on a specific domain such as the natural sciences, engineering, social sciences and humanities. Typically, these programs also have a single coordinator or director. Many also have dedicated librarians that contribute GIS and spatial analysis skills. It is also common for more traditional liaison librarians to contribute as well. Of the data management service programs at comparable institutions, the most reputable appear to take the more centralized approach.

The services catalog outlined above represents a substantial increase in services over what is currently provided. Penn State currently has several librarians who could fill these rolls, such as the Science Data Librarian and Geospatial Services Librarian. Penn State also currently offers data services in the form of data management consultation, data archiving and statistical and geospatial help. However, access is currently somewhat fragmented from the user’s perspective. We identify four core areas of service we currently provide or plan to provide, Research Data Management (RDM), Data Analysis provided through the Data Learning Center (DLC), Geospatial Analysis provided through the new Maps and Geospatial Information unit, and Digital Humanities provide through the Center for Humanities Informatics (CHI). Our goal is to maintain and augment the current level of service provided and facilitate collaboration and communication among these core service providers to further enhance our level of service.

We recommend developing a more centralized approach to developing and offering research data services at Penn State University Libraries. Given the size of the institution and its associated research community we believe this is the best way to fulfil the needs. The current RDM, DLC and GIS programs can be brought together as a single unit. This would facilitate communication and simplify access by the users. Initially, we can use our current staffing to start filling the needs, with a single coordinator to facilitate the services across the groups. In this scenario, the DLC would retain its current geospatial consultant and hire a vacant statistical analyst position. Geospatial services are provided by the new maps and geospatial information unit.

RDM services are currently provided to the sciences and engineering by the Science Data Librarian. However, data management varies somewhat across disciplines and often requires domain specific knowledge to provide the best service possible. We recommend additional strategic hires as follows to ensure that in the future the following broad subject domains are adequately served.

* Data Librarian - Engineering (Faculty)
* Data Librarian - Social Sciences (Faculty)
* Data Librarian - Arts and Humanities (Faculty)
* Data Curation Specialist (Staff)

These positions would reside in the Research Data Services Unit. These domains of the data librarians correspond roughly to our branch campuses at UP and similar strategies have been used successfully at other institutions. A social sciences data librarian is currently part of the strategic hiring plan. Until these hires are made, services can be provided by the current Science Data Librarian in collaboration with the CHI and liaison librarians from the respective disciplines.

We recommend hiring a data curation specialist who will work with the Digital Preservation Librarian, the research data librarians discussed above and the data repository managers to ensure that that the university’s full scholarly output, including research data, software, images etc. are adequately documented to help promote reuse and replicability.

While the personnel will likely all reside at the University Park campus, measures should be taken to ensure that other Penn State campus are aware of the services provided. In addition, librarians on all campuses can be trained to offer some level of service to their clientele.

If collecting the RDM, DLC and GIS services together in one group is not practical, then keeping each of the service providers separate is an alternative. We still recommend a single coordinator to facilitate communication and enhance collaboration between the groups. We also would recommend the same hiring, where the data librarians and the data curation specialist would all reside in the RDM group.

Regarding hardware, software, storage capacity and other basic needs of designing research data services, it is recommended to first determine how many existing PSU resources could be used before suggesting UL purchase all related infrastructure. The libraries already partner with ACI for storage, and SAS for software development. Given the potential scale and complexity of many research projects, we see these partnerships and others as being potentially very productive. Some of the shared resources could also be obtained through partnerships with collaborators outside of UL (see Partnership section).

Researchers use a wide array of software and other technologies. Purchasing and maintaining licenses for all of them is probably not feasible, however there is efficiency in having the library purchase and maintain some more commonly requested software and providing public access so that these resources do not have to be duplicated in every research lab on campus. In addition, having library personnel who are skilled in all of them is probably not realistic. However, the library can play an important role. By focusing on widely used tools, and promoting certain products that are known to conform to best practices and are open source, the libraries can both better manage its own resources, but also promote better more efficient practices among researchers. Also, we may be able to partner with faculty and staff in other departments who have skills we do not have.

A final consideration is the space required to offer research data services. Since space is a major challenge at PSU and many other institutions, it may be difficult to have an official office or meeting room. If possible, a dedicated room in a centrally located space with the appropriate hardware and software equipment would be ideal, where consultations and workshops could be held. Centralized space would offer the opportunity to explore additional services such as a data visualization lab, if there was demonstrated demand. However, most of the services described in our services catalog can also be offered virtually, although some services such as GIS require access to specialized tools. The Data Services web presence is of great importance and its public facing web pages should be carefully constructed to maximize ease of use, and provide our users with an easy route to finding help as needed.

## Marketing

Marketing of research data services, subject liaison librarians at University Park should play a key role. Promoting resources at faculty meetings and other departmental or institutional venues along with departmental chairs would be a good way of marketing services. It may be useful to have small handouts to leave at venues as well as in the various libraries across campus. The content from the handouts could be embedded into an email and shared across departments and institutes through liaison librarians. A consult and involvement with the UL Public Relations and Marketing group (Jill Shockey) would be very helpful in streamlining marketing approaches as well as designing of materials.

Presenting the services to librarians needs to be a part of the marketing strategy so that librarians are aware of current offerings and can consider best options for sharing that information with their various constituencies.

An organized and easily navigable web presence is an essential component of the marketing strategy. Research Data Services is a page on the current UL website and showcases current offerings and contact information, although this page currently requires substantial refinement. Any changes to services should be adjusted in a timely manner, as this will be the main source of information for users.

The Data Learning Center has had good success with their marketing and similar models should be employed for the library’s greater research data management services.

## Partners

Research data is both complex and large and requires both skills and resources that cannot be met by the library alone. Fortunately, there are several current and potential partners the libraries can work with to ensure that faculty and students needs are fully met.

The libraries are already working with the Institute for Cyberscience and Advanced Cyberinfrastructure (ICS-ACI) to ensure that we have adequate storage for our repositories. Other institutions have similar partnerships and this is beneficial in that the two entity’s skills complement each other very well. The ICS-ACI is also offering cloud computing and storage services for researchers. The libraries would be a natural partner in developing strategies to ensure that the work that occurs on these cloud services are easily managed over the data life cycle and eventually archived appropriately.

Research data and computing is a complex topic. Penn State is currently taking a more centralized approach to governing research data computing in the form of the Research Computing and Cyberinfrastructure council. It is composed of an advisory group and several working groups composed of IT professionals, faculty and university administration to focus on easing various points of friction and to help provide better services in the fields of data services and cyberinfrastructure. There PSU libraries are currently represented both on the advisory group and on many of the working groups to help guide progress.

The Data Commons is a data repository at PSU that was created to fill needs of researchers primarily in the Energy and Environmental Sciences for depositing data. The libraries are currently partnering with Data Commons to help provide a more seamless and coherent suite of services, and to build upon each other’s strengths. For instance, the libraries are working with Data Commons to help them generate DOIs for their data objects, and to provide better preservation of their content. In turn, Data Commons and their mediated deposit process remains the preferred repository for large and/or complex datasets.

The College of Information and Science Technology (IST) is a potential partner we can work with. The college has recently created the Data Sciences program which will produce data scientists with skills in data mining, data analysis and visualization. The libraries could partner with IST to create workshops and classes that help their students, as well as students from the broader institution, learn skills to better manage and document their data. Several peer institutions have piloted similar programs with success.

## Conclusion

Research data services are increasingly offered by libraries at peer institutions. In addition, PSU administration is interested in providing better data services to better serve researchers and make them more competitive for funding. The services and organization and services outlined in this report should help the PSU Libraries move forward by using existing resources in a more coordinated and efficient manner. Additional hires are suggested to augment our existing skills. Also, the organization and staffing recommended are consistent with that of many peer institutions. By collaborating with other partners on campus we can further enhance our services by leveraging the skills and resources we have and augmenting them with other campus partners.

## Literature Cited

Tenopir C, Dalton ED, Allard S, Frame M, Pjesivac I, Birch B, et al. (2015) Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. PLoS ONE 10(8): e0134826. doi:10.1371/journal.pone.0134826

Carol Tenopir, Robert J. Sandusky, Suzie Allard, Ben Birch, Research data management services in academic research libraries and perceptions of librarians, Library & Information Science Research, Volume 36, Issue 2, April 2014, Pages 84-90, ISSN 0740-8188, <http://dx.doi.org/10.1016/j.lisr.2013.11.003>.

## Committee

Robert Olendorf (Chair)

Stephen Woods

Nathan Piekielek

Robyn Reed

Rama Adithya V.