



Recommendations of the Campus Data Management and Curation Advisory Committee

November, 2011

Jeremy Frumkin and Anna Rosa Lampis, Co-Chairs

Bryan Heidorn

Chris Kollen

Dan Malone

Stuart Marsh

Derek Masseth

Nirav Merchant

Elizabeth Taylor

Committee Charge

In early 2011 the National Science Foundation (NSF) mandated that the submission of all new grant proposals include a two-page data management plan to address the following five areas:

- The types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;
- The standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies);
- Policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
- Policies and provisions for re-use, re-distribution, and the production of derivatives; and
- Plans for archiving data, samples, and other research products, and for preservation of access to them.

Fully expecting that other government funding agencies might soon have similar requirements, the Office of the Vice-President for Research (VPR) and the University Libraries charged the Data Management and Curation Advisory Committee to provide input on the following points:

1. The desirability of a University-wide strategy for data management in support of research activities
2. Identification of needed support mechanisms such as services, expertise, and technology (short term and long term)
3. Identification of the needed cyber infrastructure (short term and long term)
4. Organizational models and distribution of responsibility (short term and long term)

1. The Desirability of a University-wide Strategy for the Support of Data Management for Research Activities

1a. Recommendation: The committee recommends that the University of Arizona pursue an institutional strategy for the support of data management for research activities.

The University of Arizona currently has no formal strategy or institution-wide approach to the support of data management. This gap in support makes it difficult for investigators to:

- Find information regarding data management requirements by funding agency
- Take advantage of available campus expertise during the development of data management plans
- Have access to reasonably priced storage and preservation services for research data

A lack of institutional coordination has the potential for a long-term negative impact on the success of grant proposals to NSF, National Institutes of Health (NIH), and other funding agencies that require data management plans; as agencies begin to include consideration of the strength of data management plans in their evaluation of research proposals, the ability for the University of Arizona to demonstrate institutional support in data management plans will become increasingly important.

It should be noted that while there is no current institutional coordination, the University does provide a number of resources in support of research data management. These can and should be leveraged going forward:

- The University Libraries has created the new position of Data Management Librarian for the purpose of designing data management support services from the Libraries. These services include
 - expertise and consulting with metadata
 - development of data management plans
 - campus repository service
 - development of a data management portal to act as a central hub and starting point for researchers needing data management support
- Through the new High Performance Computing (HPC)/ High Throughput Computing (HTC) facilities, University Information Technology Services (UITS) is building out storage capacity and services to support research data storage (in addition to already-in-place HPC/HTC services)
- The University Information Security Office (ISO) provides expertise in data security
- Arizona Research Labs (ARL) provides data management support for projects within its domain
- Other units provide some support for research data management; this varies by unit, depending on the resources and expertise in a particular unit
- Individual researchers in some units have substantial expertise that could be shared with more junior faculty and students

A coordinated, campus-wide strategy in support of data management for research would benefit investigators across the institution, regardless of specific unit support. Additionally:

- Through economies of scale and synergies, smaller research efforts would benefit from support that they could otherwise not easily afford.
- Larger efforts could benefit from greater cost efficiencies and more robust support.
- It could provide the University a competitive advantage with grants where a strong data management plan is an essential factor for a successful proposal. For instance, in data management plans that require a level of data preservation and access beyond the funded term of the project, a University policy demonstrating institutional commitment for the support of data management would provide a stronger incentive than a case where there was not institutional support for such activity.

2. Identification of needed support mechanisms such as services, expertise, and technology

2a. Recommendation: The University Libraries, in collaboration with the Office of Research, should be supported to provide an integrated point-of-service with information and help to researchers about data management and the creation of data management plans.

It is critical that researchers have a starting point when developing data management plans and pursuing data management activities. Point-of-service support for data management, curation and planning should be coordinated by the University Libraries and specifically by a Data Curation Librarian. Components of the service should include:

- A robust website with:
 - Information about NSF and other funding agencies' data management requirements
 - Templates or guides for faculty to use as they are developing their data management plans. These would include the NSF guideline and suggestions for answering each section. Any standard boilerplate text developed should be incorporated into the template.
 - Listings of available data storage options and other available resources.
 - Available tutorials and/or training
 - Tools to help researchers calculate costs and budgets for their data management plans should also be made available on the web site.
 - Additional resources and/or information as they become available
- Assistance in addressing researchers' questions and linking them to the appropriate resource or unit on campus.
- Online tutorials or training sessions to provide basic information about creating data management plans and best practices for managing research data.

The University Libraries has developed the first phase of a data management web site for the campus. This website can be accessed at <http://data.library.arizona.edu>. Initial information available from the site is organized into the following categories:

- Data Management Plans
- Sharing Research Data
- Metadata Standards
- Data Management Tips, Costs and Budgets
- Campus Resources
- Library and University Data Services

2b. Recommendation: The University should pursue a coordinated communication plan to inform faculty and researchers of data management and data management planning support.

It is important that researchers know of the available support for data management and data management planning. The VPR, University Libraries should coordinate communications activities to inform all faculty, and especially new faculty, of available resources and support. These two units routinely work with faculty across the institution, and can market research data management support resources through their own services and information sharing. In particular the VPR's Office and the Library can provide information on their main websites to link researchers to the Data Management web site.

It is also acknowledged that campus business managers have a key role in their support during proposal preparation and should be periodically provided with information regarding available data management services and resources. The VPR office could provide this service through its Sponsored Projects Office.

3. Identification of needed cyber infrastructure (short term and long term)

3a. Recommendation: UITS should be supported to work with other campus units to collaboratively provide a cost-competitive multi-tiered, multi-protocol storage service that can be utilized to meet the short and long term data management needs of the institution. This business model must enable researchers to easily budget their data management costs should they choose to utilize the services provided by UITS.

One of the most immediate and clear research data management needs to be supported is the storage of, preservation of, and access to (both internal and external) to research data. In the current environment, there are varying levels of technologic support for these activities. On one end of the scale, the Arizona Research Labs provides this support directly to their researchers, while on the other end, some researchers have no support at all. While large research efforts may afford the costs of self-funding their data management infrastructure, it is likely that a great deal of research efforts would benefit from a cost-competitive, centrally provided data management service option to support their needs.

UITS is currently improving and expanding on its HPC/HTC offerings but presently there is not a specific set of centrally provided research data management support solutions available to all researchers. For the near future, UITS should leverage the HPC/HTC infrastructure to provide research data storage and access services.

3b. Recommendation: UITS, in collaboration with the University Libraries and the VPR's Office, should work to devise an overall long-term strategy for the support of data storage, data access, and data preservation in support of UA's research needs.

With regard to a long-term approach to providing this type of support, UITS, working with the University Libraries and the VPR's Office, should explore various collaborative opportunities that would enable richer and more robust infrastructures to support research data management. There are and will continue to be inter-institutional collaborative opportunities (such as DataOne [<https://www.dataone.org/>]) that could help the University of Arizona extend its abilities to support data management through technology; there is also the opportunity to explore commodity cloud-based services (such as Amazon's AWS) as potential technology solutions. Regardless of whether local infrastructure, inter-institutional collaborations, or cloud-based commodity services are utilized, UITS should be the point of service for researchers, thereby simplifying their task in obtaining the proper technology to support their data management needs.

4. Organizational models and distribution of responsibility (short term and long term)

4a. Recommendation: A Data Management and Curation Committee should be created and charged to monitor institutional support for data management and provide at least annual reports and recommendations on the state of the institutional data management and curation efforts and improvement recommendations to the Senior Vice President for Research, the Dean of the Libraries, and the University CIO.

The area of research data management and curation is a rapidly developing field. It is important that the University look at a sustained strategic approach to support it and monitor trends and emerging technologies to support activities and champion investment. While these committee recommendations, if accepted, will provide an initial strategic direction, it is also critical that the progress be monitored and adjusted over time. We recommend the creation of a Data Management and Curation Committee to regularly monitor the level and quality of data management and curation across the institution, and make at least annual recommendations regarding changes to the institutional strategy and support.

We recommend that the Data Management Committee consist of 4-6 faculty members who represent a diverse range of research areas, expertise, and experience with data management.

Committee members could be appointed for 3 years, with their terms staggered so that one-third of the committee changes each year (initial appointments will have 1/3 of the members appointed for the full three years, 1/3 appointed for 2 years, and the final 1/3 appointed for one year)

Additionally, we recommend the committee also include three ex officio members – one each from the University Libraries, the VPR's Office, and UITS.

4b. Recommendation: The University should foster internal and external collaborative opportunities aimed at continuously improving research data management

Opportunities to leverage knowledge, expertise, and resources both within and outside the institution are currently available and others will present themselves through time. As an example, currently there are campus units that provide research data management support; campus faculty and researchers with relevant experience and expertise that if shared would be of value to other researchers developing data management plans; and finally, there are inter-institutional resources and services that may be available to our researchers.

Faculty Peer Support:

- A listing of relevant data projects could be compiled through time and made available through the Data Management web site.
- The University Libraries should provide a dynamic discussion system within the Data Management web site (<http://www.coordino.com/> is an example of such a system)

Intra-Institutional Support:

- The Data Management Committee could bring together various campus units that provide data management support for the purpose of communication and coordination.
- A list of units that support data management activities, and details of each unit's support, could be compiled and made available on Data Management web site.

Inter-institutional Support:

- The Data Management Committee could explore opportunities to engage inter-institutional efforts for the purpose of expanding the UA's ability to support research data management activities. Some of the effort, but not limited to, that the committee might explore are:
 - Data Conservancy
 - DataOne
 - Vivo