Developing an Institutional Research Data Management Plan Service

January 2013
Authors and Contributors

Special thanks go to the following ACTI Data Management Working Group authors and contributors to this report.

Primary Authors

Michael Fary
Enterprise Data Architect
University of Chicago

Kim Owen
Advanced Applications Outreach Coordinator
North Dakota State University

Contributors

Erik Brisson
Associate Director
Boston University

Clifford Lynch
Executive Director
Coalition for Networked Information

Judith B. Caruso, ACTI-DM Chair
Director of Policy and Planning
University of Wisconsin-Madison

Mark Ratliff
Digital Repository Architect
Princeton University

Thomas Dopirak
Senior Consulting Architect
Carnegie Mellon University

Jina Choi Wakimoto
Faculty Director, Metadata Services
University of Colorado, Boulder

Curtis W. Hillegas
Director of Research Computing
Princeton University

Steve Wilcox
Storage Administrator
University of Wisconsin-Madison

© 2012 EDUCAUSE. The text of the report is licensed under Creative Commons by-nc-nd 3.0.
# Table of Contents

1. Introduction ........................................................................................................................................ 1

2. What Should Be Included in the DMP .................................................................................................. 3
   2.1. Type of Data ................................................................................................................................... 5
   2.2. Data Standards ............................................................................................................................... 5
   2.3. Project Storage: Provisions for Archiving and Preservation ......................................................... 6
   2.4. Access Policies and Provisions for Reuse of Data ........................................................................... 6
   2.5. Long-Term Plans for Transition or Termination of Data ................................................................. 6

3. Developing a DMP Service at Your Institution .................................................................................... 7
   3.1. Organizing, Funding, and Delivery ................................................................................................. 7
       3.1.1. Organizing: Finding a “Home” for the DMP Service ................................................................. 8
       3.1.2. Funding Model ....................................................................................................................... 8
       3.1.3. Service Delivery: Instructions and Guidance ......................................................................... 9
       3.1.4. Analyzing Impact on Campus Cyberinfrastructure ............................................................... 10
       3.1.5. Institutional Vetting of the Data Management Plan ............................................................... 11
   3.2. Researcher Resources ..................................................................................................................... 12
       3.2.1. Storage Options ...................................................................................................................... 12
       3.2.2. Productized DMP Tools vs. In-House Templates ................................................................. 13
       3.2.3. Research Administration Management System Software .................................................. 14
       3.2.4. Use of National, International, and/or Discipline-Specific Repositories .................................. 14
   3.3. Consulting ...................................................................................................................................... 15

4. Skill Sets Required for a DMP Consulting Service ............................................................................. 15
   4.1. Skills Relevant to All Types of Data ............................................................................................... 15
   4.2. Special Skills for Different Disciplines ....................................................................................... 17

5. Conclusion .......................................................................................................................................... 18

Appendix A: The Data Lifecycle .............................................................................................................. 19

Appendix B: A Sampling of DMP Support and Services ........................................................................ 22

Appendix C: ACTI-DM Working Group Members .............................................................................. 24
Executive Summary

Institutions and researchers have worked together for years on streamlining and improving campus processes to sustain institutional research at all levels. The challenges have been—and remain—many. As the data generated by research have evolved to primarily electronic format, so have the requirements for how that data is ingested, managed, and archived by each researcher and by each institution. This white paper provides insights on developing services to assist with creating research data management plans at higher education institutions, based on a broad sampling of trends in these services at institutions across the United States and internationally. From that research, a short list of key tasks have emerged as central to approaching the development of a successful data management plan (DMP) service. Institutions should consider the following as they plan how to best support researchers in preparation of their proposed DMP:

- Identify a model for local administration of research data management plan services.
- Provide resources that can be accessed conveniently by researchers during the proposal development process.
- Designate one or more dedicated staff to be available for a range of consulting needs.

This white paper includes the following sections:

- What Should Be Included in the DMP
- Developing a DMP Service at Your Institution
- Skill Sets Required to support a DMP Consulting Service

It should be noted that the focus of this document on providing a DMP service is but one component of a much larger, more formalized process at most research institutions to develop and maintain a research data management infrastructure. Additional critical topics will be covered in the related ACTI Data Management Working Group paper on developing a research data infrastructure and will include such areas as technical infrastructure, governance, and policy of the research institution. The paper is currently in draft form and is expected to be published in 2013.

1. Introduction

Institutions and researchers have worked together for years on streamlining and improving campus processes to sustain institutional research at all levels. This support has ranged from assistance in initial planning and submission of grant proposals to implementation and follow-through of those same grant activities. The challenges have been—and remain—many. As the data generated by research have evolved to primarily electronic format, so have the requirements for how that data is ingested, managed, and archived by each researcher and by
each institution. In addition, a critical compliance effort faces higher education institutions now as a growing number of federal funding agencies require that a data management plan (DMP) accompany each grant proposal submission. These compliance efforts that have been identified by the federal government as a way to ensure that research activities and results funded with public dollars are available to the public for future analysis and discovery. Indications from these agencies are that this is the first step in what will be a more comprehensive approach to research data policy.

Information technology departments on higher education campuses are keenly aware of their role in strengthening campus services to adequately support the various stages of research activities and, in particular, how the resulting research data is managed throughout its life. To successfully address these evolving needs, a formal research DMP service can be part of a larger data-lifecycle management process. A dedicated service specific to this need allows for timely communication and planning on how to best meet the needs of each research project, while also fulfilling requirements imposed by external stakeholders, like that of the federal agency DMP criteria.

This white paper was developed in an effort to continue discussion on this topic by the EDUCAUSE ACTI Data Management (ACTI-DM) Working Group. The goal of this paper is to provide insights on developing data management planning services at research institutions, based on a broad sampling of trends in these services at institutions across the United States and internationally. The evolution of agency requirements, the critical issues related to today’s academic and research data, and responses to these concerns by institutions across the United States and around the world make the contents of this document extremely time sensitive. Also, this document, because of the current urgency associated with creating DMPs for federal grant proposals, focuses primarily on requirements of these agencies. The ACTI-DM working group recognizes that a solid data management plan is of value for all research projects, regardless of funding agency, and expects that many of the ideas presented in this document would also assist in the development of those plans.

ACTI-DM believes that a formal, institutionally supported service will not only result in more accurate and comprehensive DMPs but will also generate a broader knowledge-base and community-minded environment in support of this service within the institution. This document is intended for researchers, library staff, IT staff, senior faculty, directors of research computing, CIOs, and others involved in assisting researchers develop DMPs for their grant proposals.

---

1 The National Science Foundation requirement, discussed later in this document, is perhaps the most prominent example of this trend. This project is not alone, however. For a complete list, see the California Digital Library guide, Federal Funding Agencies: Data Management and Sharing Policies, available at http://www.cdlib.org/services/uc3/dmp/fundguide.html.
It should be noted that the focus of this document on providing a DMP service is but one component of a much larger, more formalized process by research institutions to develop and maintain a research data management infrastructure. Additional critical topics will be covered in the related ACTI-DM paper on developing a research data infrastructure and will include such areas as technical infrastructure, governance, and policy of the research institution. This paper is currently in draft form and is expected to be published in 2013.

As faculty, staff, and students expand their research activities, there is a corresponding increase in the impact on campus services and cyberinfrastructure. For many higher education institutions, research activities grow ever more critical to their university’s identity, serving to attract the best faculty, staff, and students. Yet as higher education continues along this trajectory, campuses find themselves scrambling to coordinate and prepare their cyberinfrastructure and services in order to best support even current levels of research, let alone prepare for future needs.

Sally Jackson and Mike Grady of the University of Illinois at Urbana-Champaign further emphasize this urgency:

“[E]veryone acknowledges that an effective research cyberinfrastructure depends on coordination between campus resources and above-campus resources. What has not been made explicit as yet is the degree to which it also depends on coordination extended all the way to ‘below-campus’ resources—resources acquired for individual research projects through autonomous spending decisions of PIs [principal investigators] and autonomous problem-solving by university IT staff.”

To strengthen support for campus research activity and streamline the processes involved, the challenge is to create and manage the process of DMP development, both as part of the funding proposal and the overall research data lifecycle management process. This service will serve to strengthen the institution’s ability to address current and planned research activities, while addressing the needs for efficient and timely support of data management. Such an endeavor requires the coordinated efforts, knowledge, and experience of many constituents of the institution.

2. What Should Be Included in the DMP

In order to recommend institutional data management infrastructure and define research data services an institution might pursue, it is essential to understand the data lifecycle and how a data management plan provides value to this process.

---

3 Sally Jackson and Mike Grady, Introducing “Cyberinfrastructure Impact Analysis” as a Tool for Managing a Complex Sociotechnical System (University of Illinois at Urbana-Champaign, 2010).
The model illustrated below represents common data lifecycle stages assembled from examination of a number of Research Data Lifecycle models that have emerged recently (see Appendix A: ). The models studied present common data lifecycle stages from which data management needs can be determined. These common stages could be described as **data creation (conceptualization)**, **data collection and description**, **data storage, archiving and preservation**, **data access, discovery and analysis**, and **data reuse and transformation**.

![Figure 1: Common Data Lifecycle Stages](image)

The current move by federal agencies to require DMPs represents a continuing effort to address evolving standards on data lifecycle management. One factor supporting this effort is the mandate by the Office of Management and Budget under the Freedom of Information Act that aims to provide public access to research data. The mandate, last amended in 1999, articulates requirements for record retention.

---

Although the federal agencies identified in this discussion began moving toward compliance shortly after this mandate was instituted, precise definitions and criteria are still nebulous for the research institution. It must be emphasized that the requirements and resulting performance metrics are in the early stages of development. Likewise, the emergence of best practices originating from institutions in both the United States and among our international colleagues will continue to increase as this initiative matures.

Federal agencies are in the spotlight for being the first to include DMPs as a required component in a grant proposal. And because of the NSF’s prominent position within federal agencies that provide funding for research, many other federal agencies are aligning themselves to the DMP guidelines outlined by NSF. Additionally, although the focus of this discussion is on federal agencies, it should be noted that similar criteria might also be adopted by other state and local funding agencies in the near future.

The NSF has indicated that for proposals submitted to one of its directorates, the DMP should be “no more than two pages.” This presents a challenge in terms of balancing brevity with sufficient detail to satisfy the funding agency. As the process, context, and content of DMPs evolve, instructions and guidelines will be clarified. At the time of this writing, NSF requirements range from a minimum of a single sentence (primarily used in situations where a DMP is not needed) to the two-page maximum.

NSF has initially proposed a short list of recommended elements that should be included in a DMP, indicating that more details in the plans will be required in the future. Currently, these five elements are identified as data types, standards, project storage, access policies, and long-term plans. The following paragraphs discuss each element and provide relevant examples.

### 2.1. Type of Data

“Data” are primarily determined by the community of interest and may include, but are not limited to, datasets, publications, samples, physical collections, software, and models. It is recommended that DMPs include a description of data types and how that data will be created or captured.

### 2.2. Data Standards

Data management plans should specify any data standards that need to be applied for format, metadata, etc. Metadata should identify the file formats to be used for the data in question and describe any contextual details dictating metadata to be associated with that data. The ultimate goal is for data formats to be interoperable and mobile across systems (i.e., between institutional and domain repositories) to optimize discoverability and accessibility. Data should be

---

trustworthy; at its core a common framework is needed for enabling data to be discoverable and mobile. Depending on where the data will eventually reside, specific metadata standards may be required.

### 2.3. Project Storage: Provisions for Archiving and Preservation

Plans should describe storage preparation and location of the data while the research is in progress. PIs may plan to use storage available through central and discipline-specific repositories and data centers. Descriptions should include any transformations that will be necessary to prepare data for preservation and/or data sharing (e.g., data cleaning, normalization, or removing personally identifying information where appropriate).

### 2.4. Access Policies and Provisions for Reuse of Data

Critical to the open-access discussion currently gaining momentum will be a description of how data will be accessed and notes on any special resources that are needed, such as equipment, systems, expertise, and so forth. This information should also include access timelines, procedures, embargo periods (e.g., any restrictions or delays on data sharing needed to protect intellectual property), and technical mechanisms necessary for dissemination. PIs are encouraged to consider any ethical and privacy issues related to sharing the data and how those will be resolved if the data are shared (e.g., removing personally identifying information, etc.).

As indicated above, it must be noted that the larger global-scale discussion surrounding open access to publicly funded research at this time continues to escalate. If current trends continue, collective decisions by higher education around the world on how and where research findings are published and accessed will impact the processes and support services discussed here.

### 2.5. Long-Term Plans for Transition or Termination of Data

Similar to the project storage and access policies, plans for the transition or termination of data should further clarify details for the long-term strategy for maintaining, preserving, archiving, and accessing the data generated from research. Procedures should include descriptions for long-term data storage, preservation, and backup.

Plans should address how long the data will be kept beyond the life of the project. An assumed precept at this time suggests that the minimum data retention period for research data is three years after conclusion of the grant award or three years after the data are released.

---


to the public, whichever is later.

In the planning process for the transition of data to be shared with off-campus communities and to be made more public, thought should be given to any type of confidential data that might be part of the offering. Intellectual property issues may arise that an institution should consider. Personally identifiable information (PII) that may reside in the data should be put through an anonymization process, whereby any PII data are removed and substituted with identifiers that allow for indexing the data without pointing to specific individuals. This is an issue that should not have to be solved by each researcher. A DMP service offered by an institution can provide tools that would render the data anonymous and therefore freely shareable.

3. Developing a DMP Service at Your Institution

An institution that is contemplating a service to assist researchers in developing their DMPs should consider a number of factors. These factors include how the service is organized, how it should be funded, and what representatives from the institution should be included or consulted in the creation and vetting process of the DMP. In addition, the technical and consultative resources that are available to researchers for inclusion in their DMPs should be emphasized to ensure the best utilization across the institution. The following sections look more closely at these factors and what you will need to consider when developing your own DMP service.

3.1. Organizing, Funding, and Delivery

The organization of a DMP service will vary by institution and will be shaped by a number of factors, including institutional culture, organizational makeup, and geographic dispersion. Culture will provide a perspective that takes into account centralized and decentralized planning participants. The organizational makeup will help to identify the appropriate groups that influence the planning process, and geographic dispersion will require a model that allows for multi-campus institutions. Funding for the service will be driven by the organizational model insofar as budgets are most often linked to academic and administrative areas. Those areas that offer resources to the service will be most likely required to budget accordingly. An alternative to this approach would be for the institution to provide a central funding model for the resources required, wherever they reside. Delivery will be guided by the stakeholders who determine the workflow of the review, assessment, and approval of the plan and by any software tools utilized to gather the required documents for submission to the funding agency.

A particularly challenging decision embedded in this process is to identify the best time within the proposal preparation sequence to interject attention to and development of the DMP itself. The days and hours leading up to grant submission are filled with all the final edits, fact checking, and various other tasks of high priority. Intervening with yet one more item added to
the list is not going to be met with broad support, at least in the early stages. At the time of this writing, there is indication that campuses are beginning to request deadlines for DMP consulting and vetting within a set window prior to grant submission. Ultimately, what emerges as a widely accepted and cogent practice for DMP development is yet to be seen.

3.1.1. Organizing: Finding a “Home” for the DMP Service

Identifying a model for local administration of DMP services is likely to be a primary objective for any campus. Developing the model that best meets the needs of the local campus research environment requires the engagement of departments across campus. From an institutional standpoint, this service can be viewed as a core infrastructure activity in support of researchers. For the researchers, it provides valuable expertise needed to prepare the DMP component of the grant proposal, as well as critical insight to existing and planned institutional resources that can be leveraged to make the most of grant dollars received.

As indicated earlier, because best practices in the administration of this process are still emerging, very few examples of widely accepted and well-tested models exist. In this early stage, investigation by ACTI-DM has found that campuses across the United States appear to be gravitating toward two preferred models for administration of research data management plan services. These models are described below:

1. Embedded Small Group: A small group of designated staff are embedded in an existing department whose activities are closely aligned and/or impacted by these efforts (i.e., library, research, IT, or in some cases an independent office supported by all three of these and possibly others); additionally, an advisory committee may be named whose charge is to provide oversight and guidance to the staff in developing protocols that best serve a large spectrum of departments and disciplines as well as provide guidance embedded in specific discipline topics.

2. Advisory Committee: An advisory committee is selected and charged with the responsibility of providing guidance and assistance to individual faculty and departments for a collection of tasks related to data management (e.g., developing DMPs for grant proposals including discipline-specific guidance, providing DMP review services in the pre-proposal phase, identifying IT costs related to proposed research projects, etc.). These committees typically consist of cross-departmental representatives including but not limited to library, research, IT, faculty, sponsored programs, etc.

3.1.2. Funding Model

A funding source must be identified to cover staffing and resources needed for this process to function efficiently and effectively. Roles and services identified in the above steps impact budget needs for this initiative. Institutions might choose to begin by building support and
refining the details of this process and only addressing the needs of federally funded grant proposals, adding processes for other state and local funding agency proposals as appropriate. Proper funding of the consulting activities that develop the DMPs must become a foundational process of the institution. If these activities are to be “backhanded” by the required staff resources, other priorities will take precedence and the DMPs will suffer. The consulting activities and related resources must be considered a part of the infrastructure of the institution, much as any other basic service that provides support to the research activities of the institution.

3.1.3. Service Delivery: Instructions and Guidance

Each institution will have its own methods and workflows for facilitating a DMP service, generating the various necessary documents for proposal submission, bringing that information together, and submitting the proposal to the granting agency.

Depending on the institution, there may be steps that include the development of the plan, routing to the appropriate reviewing and vetting parties, approval, and association of the plan with the other documents required for the proposal submission. Any process model and collection of resources needs to be customized in a fashion that works best for both the campus community and individual disciplines and that functions in a way that is the most effective and yet the least invasive to the grant pre-proposal process. There cannot be a “one size fits all” that is identical for every institution or discipline.

Various departments across the institution might be consulted or utilized in the process of creating the DMP. Each might contribute resources, expertise, or guidance in the development and preparation of the DMP. Since many of these entities will also provide ongoing support and/or resources to the research activities, they will have a vested interest in what is suggested or committed to in the DMP. Some areas of responsibility for an institution to consider as it develops its own processes include the following:

- Central IT: Central IT might provide support in the form of central storage, computing resources, networking, programming, database, security, and other computing-related resources. In addition, other forms of expertise might be provided. This could come in the form of a consultative role or actual hands-on development or administrative work in any of the aforementioned areas.
- Faculty/Deans: The deans and faculty of the respective research areas provide the background knowledge and impetus to the research proposal process. They also might contribute to the development of the DMP in the form of budget available for staffing and other resources.
• Legal Counsel: Legal counsel might provide guidance on intellectual property (including any copyright or patent sensitivities related to the data) or on legal issues related to personal information, health data, or other sensitive types of data.

• Library/Institutional Repository: The library might provide support in the form of digital librarian skills related to the data being produced. If the library is the custodian of an institutional repository, they could assist with the classification and ingestion of data into the institutional repository. They might also provide guidance to assist the researchers with identifying and classifying data for issuance to disciplinary repositories.

• Research Computing: A research computing organization can provide assistance in the development of the DMP related to high-performance computing and storage options, consortia available to the researcher that they may avail themselves of, and techniques related to data movement. They also can provide information related to costs and chargeback if that is required of the institution.

• Sponsored Programs: The sponsored programs office provides oversight to the federal grant proposal process, of which the DMP will be a component. In their coordination role, they can engage in the seamless development and submission of the required DMP, including providing PIs a reference to further supporting resources on campus and ensuring that the DMP is completed on time and is brought into the proposal process. They also ensure that any additionally required documentation is included.

• Other: There may be other areas that can assist in the development of the DMPs. Examples include department-level IT groups and affiliates of the institution that might either wholly or in part host computing or storage environments that researchers can utilize.

Individual service components need to be considered with the campus’ needs and culture. How these components are organized and delivered will vary from institution to institution.

3.1.4. Analyzing Impact on Campus Cyberinfrastructure

Another important component of a DMP service is to analyze the impact of the project and the data it generates on campus cyberinfrastructure. Many of those in research and campus technology leadership roles will nod to the practice of planning and communicating as early and as frequently in the proposal process as possible. Specific to campus cyberinfrastructure, this deliberate attention to data management within the overall research life cycle opens the door to clarifying how any proposed investment in research will affect its context.

Cyberinfrastructure impact analysis can be incorporated into many common decision workflows, including creation of new programs, construction of new facilities, establishment of new centers, or acceptance of major grants. The earlier this takes place, the better the
opportunities for the institution to control costs. For example, by routinely conducting cyberinfrastructure impact analysis on grant proposals, a campus might be able to produce lower-impact ways of accomplishing a researcher’s goal. At a minimum, this form of analysis allows greater predictability and financial preparation.

### 3.1.5. Institutional Vetting of the Data Management Plan

As DMPs are created, the vetting and/or approval processes should be considered. Both of the models described in Section 3.1.1 also speak to the logical selection for those involved in this process. Since these plans represent a commitment to current and future resources, budget, and risk, the institution should provide some method by which the plans are reviewed and verified for completeness and validity. Institutional vetting provides at least three valuable results:

1. Improved quality in grant applications, with the goal of more awards being granted due to higher quality DMPs.
2. Clarification of roles, responsibilities, and risks of the institution and the researcher.
3. Greater commitment from the institution and the researcher to what is included in the DMP.

The overall quality of the plan should be considered in light of a number of factors:

- **Risks to Institution and Researcher:** Risks related to loss of funding, future research possibilities, and researcher and institutional reputation.
- **Roles and Responsibilities of the Institution and the Researcher:** Ensure that the roles of the researcher and the institution are considered and spelled out with regard to tasks, services, budgets, and timelines and that both have agreed to their respective responsibilities.
- **Storage Requirements and Availability:** Storage of the data during and after the active research cycle, including the grant period. This includes storage sponsored by the institution and in disciplinary repositories.
- **Curation Needs:** Ensure that the data will be curated to the level that guarantees its viability.
- **Commitment of Technology Resources and Refresh Cycle:** Ensure that the data will be maintained in an environment that stays technologically current.
- **Viability of Network Capabilities:** Ensuring that data transfer needs will be met.

These factors ensure that a sufficiently thought out, vetted, and agreed upon DMP for a researcher will be the key deliverable of the DMP service. By assisting the researcher in developing the DMP, the institution and the researcher benefit from the combined knowledge and experience of the parties who bring their expertise to the process.
3.2. **Researcher Resources**

Universities continue to build substantial collections of information targeted directly at the PI/researcher to assist them in development of their required DMPs. Posted online, this information provides a convenient, rich collection of tools and resources that support various components of the DMP. PIs are able to develop a DMP in step with the development of the grant proposal itself, tapping resources as needed. In most cases, these resources are compiled in a campus DMP website that often resides in the library or research domain, with links included to numerous other support services and resources both on and off campus.

The value of communication and planning across all support service areas cannot be emphasized enough. Including resources and links to additional tools provided by various departments across campus will provide the PI with a one-stop shop that can meet all of their needs. The items listed here provide a solid foundation of tools on which to build a successful DMP service and includes a sampling of the types of additional services that a DMP service may want to provide for its researchers.

### 3.2.1. **Storage Options**

Provide a “researcher’s storage toolkit” identifying various types of storage services that are available. Resources can be delineated by stages in the research data lifecycle and may be available internal or external to the institution.

As data moves through the various stages of its lifecycle, it is maintained in storage environments that provide varying levels of accessibility to the community of parties who may be interested in access to the data. Initially, the data is contained in a private state, with only the research team having access. At some point, the data is moved to a state that allows it to be shared inter- or intra-institutionally. Finally, the data is moved to an archive state either directly from the private or from the shared environment.

![Figure 2: Storage Options](image)

---

12
Examples of storage environments that could house the data include any combination of the following:

- Central IT storage, in the form of a managed storage area network, for short- and long-term use.
- Decentralized research storage, such as that typically provided as part of a high-performance computing (HPC) environment, provides short-term storage for use by the researcher during the heavy computation portion of the research. This storage comes at a premium in terms of cost and availability and should not be expected to be available to the researcher for long-term archival use.
- Divisional/departmental storage environments will be utilized depending on the stage of their research and the types of access needed during that stage. This type of access will require some type of federated architecture with the other types of storage environments, requiring distributed access and control mechanisms, taking into account that the local storage may have to interact with authentication and authorization mechanisms provided by the institution or other consortia.
- Cloud-based environments, either commercial or public, offer storage services that researchers may take advantage of.

These storage services may be provided at no cost to the researcher, instead being funded as part of institutional research objectives. In some cases, fee-based services for some or all of the components may be in place to recover the cost of storage, backup, access, and recovery operations.

As collaborations in research grow and the ability to capture and manipulate data from a variety of personal devices increases, options for local and remote storage will need to stay in step with these changes.

### 3.2.2. Productized DMP Tools vs. In-House Templates

Many institutions have already created templates and examples that provide a general framework for what should be submitted to the NSF. These templates can be copied and used by researchers as a basis for developing and submitting the required DMPs. In addition, a resource such as the online tool created by California Digital Library can assist with the development of the DMP. Called the DMPTool, it provides a web-based environment to guide a researcher through the steps of creating a DMP.\(^8\) It offers guidance for specific directorates within the NSF, as well as NIH and several other granting foundations and services. The DMPTool also allows each institution using the tool to add its own specific material. At the

---

\(^8\) California Digital Library, DMPTool: Guidance and Resources for your Data Management Plan, [https://dmp.cdlib.org/](https://dmp.cdlib.org/).
conclusion of the development of the plan, a plain text or rich text file can be exported and uploaded to research administration management system software.

### 3.2.3. Research Administration Management System Software

Institutions may utilize some type of proposal and award management software. Examples of these types of systems include the MIT-developed Coeus\(^9\) and Huron’s Click\(^{10}\) solution. These systems may allow for development of the DMP directly in the application, or they may provide the ability to include an externally developed plan as an electronic document. These systems are comprehensive proposal preparation, submission, and tracking systems.

### 3.2.4. Use of National, International, and/or Discipline-Specific Repositories

The DMP will also need to include a final destination for data that are created as a result of the research activity. NSF has indicated that “[i]nvestigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants.”\(^{11}\) An obvious question then arises as to where there exists a suitable place to land the data. In addition to institutional repositories, data may be housed in discipline-specific repositories or in national or international repositories, which typically host large data sets.

For example, the Institute of Museum and Library Services, with additional support from the Purdue University Libraries, originally sponsored Databib, a tool for “helping identify and locate online repositories of research data.”\(^{12}\) Although specific definitions for terms such as “data” and “repositories” may be a moving target at this time, Databib promotes the theory that common definitions of “data repository” have surfaced and may include the following:

- A system that provides online access to research data
- By an entity that is committed to providing access to the data
- That serves as a primary, authoritative source for data
- With a significant collection of like data or data within the scope of a management policy.\(^{13}\)

Along with the deposit of materials and files, metadata are created and allocated in the manner consistent with the research discipline and the hosting venue. Critical issues concerning long-term storage include curation, intellectual property rights, and maintenance of anonymity as it relates to requirements of the institutional review board (IRB) review and approval. A Digital

---


Object Identifier (DOI)\textsuperscript{14} should be created to provide a persistent link to the data for the long term.

3.3. Consulting

Consulting is a critical addition to the value provided by any and all of the support services included in this effort. Designating one or more dedicated staff to be available is valuable to the researcher. Our review of campus approaches to integrating consulting as a key resource uncovered numerous examples, most with very similar responsibilities.

As evidenced both by the examples of staff positions on campuses across the country and by publications on this topic, popular staff positions include technical staff, who are tasked with creating and maintaining all technical resources (for both library and IT services), and specialized consultants, often with advanced degrees in specific disciplines. Campuses are filling this need either through paid staff positions or as critical members of the advisory councils that support this support service.

4. Skill Sets Required for a DMP Consulting Service

The DMP service should have access to or be comprised of individuals or organizations that have the necessary skills needed to prepare a complete and well vetted DMP for the researchers. The required skills may be common across all research disciplines, or be specific to certain types. That range of abilities is explored in the following sections.

4.1. Skills Relevant to All Types of Data

When considering the makeup of the consulting group, skill sets of various types of individuals should be taken into account. Individuals with knowledge of the following areas may need to be consulted when creating the DMP.

- Storage: Research data will proceed through a lifecycle from inception of the project to active research, publishing, and archiving. At the various stages in its life, different types of storage should be considered. Given the available storage options associated with high-performance computing, local and shared storage for intermediate results, publishing, and archiving, expertise on what storage is available where and how to use it would be valuable to the researcher.

- Data Migration: As the research proceeds, the data may have to move from one environment to another. Depending on the type, size, and context of the data, this may involve many types of technology. If the data resides in some type of database or other

\textsuperscript{14} The DOI system involves permanently assigning a DOI name to an object to provide a resolvable, persistent network link to current information about that object, including where the object, or information about it, can be found on the Internet. For more information, visit \url{http://en.wikipedia.org/wiki/Digital_object_identifier}.
structured technology, programming or utilities may be required to extract the data and put it into a more suitable format for archiving.

- Networking: If large volumes of data are to be moved within and across multiple research environments, significant network bandwidth may be required. Expertise in the tools and techniques that utilize the network pipes would be critical to the researcher.

- Legal: There may be a need for knowledge of governmental laws and regulations that may specify access, use, management, and retention of the data. HIPAA, FERPA, and personal privacy laws may have a direct bearing on the data. In addition to data protection and privacy laws, additional expertise may be required, such as knowledge of intellectual property laws.

- Financial: These guidelines may be particular to the grant parameters itself and housed in the grant budgeting office, but they will also need to be considered beyond the scope of the grant. Expertise on costs may come from various sources on campus, including IT (e.g., for data storage costs, networking, migration, etc.) and others. The DMP service can assist in coordinating budget information between what is needed for the grant itself and any considerations beyond the grant (e.g., if data is retained at the institution beyond what was scoped in the grant).

- Security: Access to the data may need to be controlled. In these cases, authentication and authorization mechanisms may need to be employed to ensure that the data are accessed properly. Experts are required to implement security controls, including those controls specific to cloud-based services.15

- Metadata Creation and Assignment: Basic metadata for discovery, interoperability, and provenance will be required as a minimum for deposit in institutional or disciplinary repositories.

- Scholarly Data Communications: There will undoubtedly be questions related to copyright, open data issues, policy, and a myriad of other data areas. The expertise to answer these questions, or at least point the researchers to a knowledgeable source, will be important to ensure that the DMPs are created with an assurance that data issues have been addressed.

- Preservation: Data preservation and archival expertise is necessary for effective data management after the initial research project has completed its use of the data.

---

15 Cloud strategy was identified as a Top-Ten IT issue for 2012 by EDUCAUSE members (see http://www.educause.edu/ero/article/top-ten-it-issues-2012), and a number of recent reports have been published about the particular security concerns around data that are stored in the cloud. See, for example, Cloud Computing Synopsis and Recommendations (National Institute of Standards and Technology, Special Publication 800-146, May 2012, http://www.nist.gov/itd/cloud-052912.cfm) and NSTAC Report to the President on Cloud Computing (President’s National Security Telecommunications Advisory Committee, May 2012, http://www.ncs.gov/nstac/reports/2012-05-15%20NSTAC%20Cloud%20Computing.pdf). In addition, a useful overview of the issues in this area is 7 Things You Should Know About Cloud Security (EDUCAUSE, August 2010, http://www.educause.edu/library/resources/7-things-you-should-know-about-cloud-security).
4.2. **Special Skills for Different Disciplines**

Different disciplines within the university will inherently produce different types of data, different volumes of data, and different contextually identifying information about the data. From the types of files being created to the metadata used to provide searchability and context, there will be a need to identify and utilize differing data skills. In order to provide a suitable environment for ongoing management of the research data, those skills will need to be tapped.

One challenge will be to identify individuals and/or job roles that may provide those specific skill sets. These skills may reside in the library environs, in the form of subject specialists or liaison librarians. They may exist in the graduate or postgraduate realm of research. And in some institutions they may need to be cultivated in new roles that do not yet exist.

As data management planning is considered, so too must the skills that are unique to certain disciplines. In order to create DMPs that are viable, those skills should be consulted and utilized.

Discipline-specific skills include:

- **Metadata Creation and Assignment**: Different disciplines often have unique standards for metadata creation, particularly when archiving is being considered. Knowledge of discipline-specific metadata is a factor when assisting researchers with their DMP proposals, and also when they are actually creating the metadata for a repository. The metadata that are created may be related to publications or other publishable works. Additionally, metadata may be created for data or data sets that are being placed into repositories or other types of archives.

- **Discipline-Specific Repositories**: Some higher education disciplines have repositories that are dedicated to gathering publications and data from their subject area. They also may have specific guidelines for storing the data, what metadata to use, and search criteria. When considering the long-term archival of data, these types of repositories may be a possible landing spot. Some examples include:
  - Inter-community Consortium for Political and Social Research (http://www.icpsr.umich.edu): ICPSR provides leadership and training in data access, curation, and methods of analysis for a diverse and expanding social science research community. ICPSR is part of the Institute for Social Research at the University of Michigan.
  - GenBank (http://www.ncbi.nlm.nih.gov/genbank): GenBank is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences. The GenBank database is designed to provide and encourage access within the scientific community to the most up-to-date and comprehensive DNA sequence information.
- NCAR’s Community Data Portal (http://cdp.ucar.edu/home/home.htm): The Community Data Portal (CDP) is a collection of earth science data sets from NCAR, UCAR, UOP, and participating organizations.

5. Conclusion

To support increasing pressures of efficiencies and compliance, institutions are encouraged to implement a data management planning service that supports their unique research environment. Early on, the objective of this service might focus specifically on supporting the DMP as it is required for inclusion in a research grant proposal. However, institutional leadership should remember that indicators point toward a looming larger task making DMPs part of the full research data lifecycle, including implications for the movement to open access of published research. Thus any and all guidelines provided here are time-sensitive, and careful monitoring of current trends across institutions will prove valuable as this work evolves.

Libraries and IT departments are only two of several areas in higher education institutions that play a critical role in the development of such a service. The technical infrastructure they maintain, as well as related tangential services available on that infrastructure, form the transparent foundation on which to build a set of services aimed at ingesting, maintaining, manipulating, and preserving research data. Ongoing planning and collaboration with various stakeholder groups across the institution are critical to the development of a service that best addresses the unique needs of research across disciplines and across their own campus.

Based on research focusing on trends at institutions in the United States, a short list of key tasks has emerged as central to approaching the development of a successful DMP service. From this list, the ACTI Data Management Working Group recommends institutions consider the following as they plan how to best support researchers in preparation of their proposal’s DMP.

- **Identify a model** for local administration of research data management plan services.
- **Provide resources for convenient access** to researchers during the proposal development process.
- **Designate one or more dedicated staff** to be available for a range of consulting needs further reinforces the value of this support system for the researcher.

The ideas and issues identified in this document are intended to assist institutions in organizing their data management plan services. Its goal is to motivate continued discussion and exploration by higher education institutions in their pursuit of service models that best facilitate the vision of preserved and discoverable research data.
Appendix A: The Data Lifecycle

A number of Research Data Lifecycle models have emerged in recent years. They include the Data Curation Centre (DCC) Curation Lifecycle Model,16 the Data Documentation Initiative (DDI) version 3.0 Conceptual Model,17 the Australian National Data Service (ANDS),18 the DataONE Data Life Cycle,19 the UK Data Archive Research Data Lifecycle,20 and the Capability Maturity Model for Scientific Data Management.21

The DCC model has eight actions for curation and preservation of research data, which are the execution of data management plans with administrative actions for curation.

1. Conceptualize
2. Create or Receive
3. Appraise and select
4. Ingest
5. Preservation action
6. Storage
7. Access, use, reuse
8. Transform

DDI version 3.0 Combined Lifecycle Model is more linear and geared towards social science data. It also has eight elements:

1. Study Concept
2. Data collection
3. Data Processing
4. Data Archiving
5. Data Distribution
6. Data Discovery
7. Data Analysis
8. Repurposing

---

19 The DataOne Data Life Cycle can be found on the DataOne Best Practices site at http://www.dataone.org/best-practices.
Both the Australian National Data Service (ANDS) and The DataONE Federation use verbs to describe data lifecycle models. Both have eight actions though expressed slightly differently. ANDS data sharing verbs are Create, Store, Describe, Identify, Register, Discover, Access, and Exploit. DataONE Federation uses data lifecycle verbs similar to ANDS, which are Plan, Collect, Assure, Describe, Preserve, Discover, Integrate, and Analyze.

UK Data Archive Data Lifecycle defines six stages of lifecycle and added activities of data management relating to those six stages: creating data, processing data, analyzing data, preserving data, giving access to data, and reusing data.

In the Capability Maturity Model for Scientific Data Management, key data management practices are grouped into four key process areas: data acquisition, processing, and quality assurance; data description and representation; data dissemination; and repository services/preservation. In addition to the key process areas, several practices that support the data management process are also identified. As such, this model can be regarded as a checklist of practices for data management.

The models studied above present common data lifecycle stages from which data management needs can be determined. The common stages could be described as data creation (conceptualization), data collection and description, data storage, archiving and preservation, data access, discovery and analysis, and data reuse and transformation.
Determining campus research data management needs will, in turn, help define which research data services an institution can provide. An institution should adopt its own data lifecycle and determine which services it is going to provide in each of its data lifecycle stages.
Appendix B: A Sampling of DMP Support and Services

The following is a list of websites, templates, tools, guidance documents, and services that serve as a resource to researchers and other involved in data management plans.

This is not intended to be an all-inclusive list, but rather a sampling of various types of services and support that institutions provide for their researchers and faculty to assist them in the development of data management planning activities.

California Digital Library
UC3: University of California Curation Center: Data Management Guidelines

Clemson University Libraries: Guide to Data Management
http://clemson.libguides.com/content.php?pid=175832

Cornell University: Research Data Management Service Group
https://confluence.cornell.edu/display/rdmsgweb/Home

Duke University Libraries: Data Management Guide
http://library.duke.edu/data/guides/data-management/index.html

George Mason University Libraries: Data Services
http://dataservices.gmu.edu/

Johns Hopkins University: Data Management Services
http://dmp.data.jhu.edu/

MIT Libraries: Data Management and Publishing

North Dakota State University, Research Data Working Group
Data Management Plans
http://www.ndsu.edu/research_data/data_management_plans/

Princeton University Library: NSF Data Management Plan Help
http://libguides.princeton.edu/content.php?pid=211802

Purdue University Research Repository (PURR)
https://research.hub.purdue.edu/

UC San Diego Library: Research Data Curation Services
http://libraries.ucsd.edu/services/data-curation/
University of Chicago, University Research Administration
NSF Data Management Template

University of Colorado, Boulder: Research Data Services
https://data.colorado.edu/

University of Connecticut University Libraries: Data Management
http://classguides.lib.uconn.edu/datamanagement

University of Edinburgh: Research Data Management Guidance
http://www.ed.ac.uk/schools-departments/information-services/services/research-support/data-library/research-data-mgmt

University of Hawaii at Manoa Library: Data Management Plans
http://guides.library.manoa.hawaii.edu/content.php?pid=125160

University of Michigan Library: Research Data Management and Publishing Support
http://www.lib.umich.edu/research-data-management-and-publishing-support

University of Minnesota Libraries: Managing Your Data
https://www.lib.umn.edu/datamanagement

University of North Carolina at Chapel Hill Library
Research Data Toolkit: Resources for Writing a Data Management Plan
http://www.lib.unc.edu/datamanagement/

University of Oregon Libraries: Research Data Management
http://libweb.uoregon.edu/datamanagement/

University of Oxford
University Administration and Services: Research Data Management
http://www.admin.ox.ac.uk/rdm/

University of Virginia Library: Data Management
http://www2.lib.virginia.edu/brown/data/

University of Wisconsin–Madison: Research Data Services
http://researchdata.wisc.edu/
Appendix C: ACTI-DM Working Group Members

The following were members of the ACTI Data Management Working Group at the time of publication.

Vijay K. Agarwala  
Senior Director, Research Computing and Cyberinfrastructure  
The Pennsylvania State University

Phillip Berres  
Software Engineer  
University of Southern California

Erik Brisson  
Associate Director  
Boston University

Judy Caruso, ACTI-DM Chair  
Director of Policy and Planning  
University of Wisconsin-Madison

Thomas Dopirak  
Senior Consulting Architect  
Carnegie Mellon University

Michael Fary  
Enterprise Data Architect  
University of Chicago

Curtis W. Hillegas  
Director of Research Computing  
Princeton University

Courtney Jones  
Assistant Director of Database Administration  
Kennesaw State University

Ardys Kozbial  
Technology Outreach Librarian  
University of California, San Diego

William J. Labate  
Director, Academic Technology Services  
University of California, Los Angeles

Clifford Lynch  
Executive Director  
Coalition for Networked Information

Mairéad Martin  
Senior Director, Digital Library Technologies, Information Technology Services  
The Pennsylvania State University

Donald F. McMullen  
Cyberinfrastructure Strategist  
Great Plains Network

David Minor  
Head, Data Curation Services  
San Diego Supercomputer Center

Kim Owen  
Advanced Applications Outreach Coordinator  
North Dakota State University

Mark Ratliff  
Digital Repository Architect  
Princeton University

Jina Choi Wakimoto  
Faculty Director, Metadata Services  
University of Colorado Boulder

Steve Wilcox  
Storage Admin  
University of Wisconsin-Madison