

# Complying With Funder Requirements for Making Research Data Publicly Available: Challenges and Opportunities for UAB.

Report of the RDM Working Group  
January 2017

This document presents the report and recommendations from the RDM Working Group. The Working Group was established in September 2015 to examine the issues and opportunities that arise from emerging federal policies that require the data resulting from federally funded research to be made publicly available. The recommendations focus on the needs of UAB investigators for support in complying with these policies.

## Report and Recommendations, pages 2-7:

- Introduction
- Best Practices for RDM
- Data Management Plans (DMP)
- Infrastructure Support
- Governance
- Resource Requirements
- Next Steps
- Members of the RDM Working Group

## Appendices, pages 8-16:

- A. Charge to the Working Group
- B. Defining Research Data
- C. Recommendation for Best Practices Guidelines Development
- D. Examples From Other Universities
- E. Results of the Survey of Investigators' Data Management Practices

# Complying With Funder Requirements for Making Research Data Publicly Available

## *Introduction*

In February 2013, the White House Office of Science and Technology Policy (OSTP) issued a memorandum directing “each Federal agency with over \$100 million in annual conduct of research and development expenditures to develop a plan to support increased public access to the results of research funded by the Federal Government.”<sup>1</sup> By the fall of 2015, these plans had been developed and are now in the process of being implemented. These new policies will require investigators to develop Data Management Plans when submitting grant applications, and to deposit publications resulting from federal funding and the data supporting those publications into authorized repositories. **Failure to establish appropriate mechanisms for complying with these new requirements could have substantial negative impacts on UAB's competitive position in acquiring federal research funding.**

In September 2015, a working group (referred to here as the RDM working group) was established by the Senior Vice Provost and the VP for Information Technology and charged with developing strategies for supporting investigators in responding to these requirements. This report is directed to them for review and possible action. It concludes the work of the group. (The charge to the Working Group is included here as Appendix A).

The issues and recommendations were developed over the course of several meetings and email discussions among the members of the Working Group. The issues involved were frequent topics of discussion at the monthly Data Wranglers sessions held in the Edge of Chaos and the comments and suggestions from attendees at those sessions had a significant impact on the content of the report. A survey distributed to all currently funded principal investigators in May 2016 provided additional insight into the data management issues of most concern to UAB investigators.

Since its establishment, the working group has:

- Developed a Research Data Management website for UAB which provides basic information about the funding agency requirements along with resources for compliance. The website is maintained by the Director of Digital Data Curation Strategies. <http://www.uab.edu/faculty/rdm>

---

<sup>1</sup> Office of Science and Technology Policy. Memorandum: Increasing Access to the Results of Federally Funded Scientific Research. February 22, 2013. [https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp\\_public\\_access\\_memo\\_2013.pdf](https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf)

- Developed a survey which was sent to all currently funded PIs to provide a general overview of the state of RDM at UAB. The survey highlighted several areas of immediate concern. The report of the survey is included here as Appendix E.
- Issued a recommendation to the VP/IT that a group representing the various stakeholder units be charged with developing a UAB set of Guidelines for Best Practices in Research Data Management. The recommendation is included here as Appendix C.

Support for investigators needs to be provided throughout the research data lifecycle and requires a mix of policy, services and infrastructure. More specifically, the following areas need to be addressed:

- Policies establishing best practices for research data management (RDM).
- Support for developing, implementing, and monitoring Data Management Plans.
- Infrastructure support for data security, storage, backup, archiving, and sharing.

### ***Best Practices for RDM***

UAB has no generally accepted and widely promulgated best practice standards for managing research data. There are a handful of policies that address some aspects of data management but it is left to individual investigators to develop and enforce procedures for handling the data generated by projects for which they are responsible. Consequently, much of the data generated by sponsored research lacks necessary backup and security routines and may not be organized or maintained in a form and format that will likely comply with emerging funder requirements.

Currently in place:

- [Guidelines for Data Collection, Documentation, and Storage](#) (from OVPRED) (2014)
- [Data Custodian Responsibilities](#) (from IT) (2012)
- [Data Protection and Security Policy](#) (2007)
- [UAB Records Retention Schedule](#) (section on Research Records) (2013)
- [Data Classification Requirement](#) (from IT) (Draft, Nov 2016; implementation planned for mid-2017)
- In September 2016, the working group submitted a recommendation to the VP/IT that a group with appropriate representation be charged with developing UAB-sanctioned Best Practice recommendations (see attachment C). VP/IT responded that such a group should include the Chief Data Officer

who is being recruited during the fall 2016. We anticipate that the process of developing Best Practice recommendations will commence early in 2017.

What is needed:

- Development of UAB-sanctioned Best Practices Guidelines for RDM that address security, backups, naming conventions, storage conditions, and long-term archiving/curation support.
- Development of a strategy for providing education and training throughout the university regarding these Best Practices.

### ***Data Management Plans (DMP)***

Since 2003, NIH has required the inclusion of data sharing plans in grant applications for proposals requesting more than \$500K in direct costs and for some proposals in specific areas. In the plan published in February 2015, NIH states that they will be “taking steps to ensure all NIH funded researchers develop data management plans whether they are funded by a grant, cooperative agreement, contract, or intramural funds, regardless of funding level.”<sup>2</sup> Although the implementation timetable is still unclear, it is the intention of NIH that these plans will be evaluated and scored as part of the peer review process. NSF has had a long-standing policy requiring DMPs for all grant applications. The other federal funding agencies are now developing similar requirements.<sup>3</sup> An increasing number of non-governmental funders also require DMPs of varying degrees of specificity.

Currently in place:

- A link to a UAB branded version of the DMPTool has been added to the RDM website. (The DMPTool is an online resource that provides guidance in developing DMPs. It is provided as a service of the University of California Curation Center of the California Digital Library.)
- The RDM website provides links to the DMP requirements of the various federal agencies.
- The CCTS Research Commons has agreed to provide assistance to investigators seeking help in preparing DMPs for grant applications.

---

<sup>2</sup> National Institutes of Health Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research February 2015

<https://grants.nih.gov/grants/NIH-Public-Access-Plan.pdf>

<sup>3</sup> SPARC. Browse Data Sharing Requirements by Federal Agency.

<http://researchsharing.sparcopen.org/compare?ids=6&compare=data>

What is needed:

- A systematic mechanism for providing guidance on DMPs before grant applications are submitted – investigators should have assurance that their DMPs are making the best and most cost-effective use of available UAB resources.
- A mechanism for assistance to investigators in monitoring the implementation of a DMP through the life of the project and afterwards.
- Integration of DMP development/monitoring with IRAP.

### ***Infrastructure Support***

UAB has some core facilities/resources devoted to helping investigators store and manage research data. Little is available for curating data for archiving and re-use. The resources that are available are often not well understood or well used.

Currently in place:

- There are a number of tools available on campus to support data collection, storage and analysis. For example, UAB Box for non-sensitive data storage (UAB IT), Cheaha for high-performance computing and high throughput computing paradigms (UAB ITRC), REDCap for building and managing online surveys and databases (UAB DOM) are some of the resources used by investigators across the institution.

What is needed:

- A coordinated approach toward identifying campus tools and resources available to support the storage and analysis of research data that is easily discoverable by the campus community.
- Guidance in identifying appropriate repositories for permanent deposit of research data, along with financial support, as needed.
- UAB managed data repository for data that does not have another logical “home”.
- Assistance with applying appropriate metadata in order to facilitate appropriate discovery and re-use of data in the future.

### ***Governance***

As described in a 2013 article in the *Educause Review*, developing the necessary policies, infrastructure and services requires a coordinated effort across many units

of the institution.<sup>4</sup> Effectively supporting UAB investigators and ensuring that we can comply with emerging funder and publisher requirements for access to research data will require coordinated action across many areas of the university. The appointment of a new VP for Research and the establishment of the position of Chief Data Officer can serve as catalysts for identifying an appropriate management/coordination structure.

### ***Resource Requirements***

Personnel are the primary resource requirement for addressing many of the issues identified above. Depending on the range and depth of services offered, institutions that have advanced in these areas have support personnel ranging from a single FTE to several (appendix D provides examples of what some other institutions are doing). Frequently these are located in the library but may be distributed across the institution depending on the expertise and training required. Identifying the specific resource needs for UAB will require identifying the specific priorities and timeline for addressing the issues raised in this report.

In addition to personnel, costs for developing/maintaining a data repository can vary widely depending on how it is approached. Options that could be explored at UAB include:

- UAB Libraries currently support a ContentDM site to support the UAB Digital Collections. This could be configured/enhanced to house small datasets.
- UAB Libraries licenses Rosetta as part of its suite of library infrastructure tools. Rosetta was developed as an archive for digital assets and is not currently configured to act as a data repository, but the parent company has indicated their interest in moving its development in that direction.
- figshare, a cloud-based data repository system available from the same company that UAB licenses the Faculty Profiles system from has an institutional offering.
- Digital Commons, from bepress, is another cloud-based system that is used as a data and document repository by many institutions world-wide.

---

<sup>4</sup> Erway R. Starting the Conversation: University-wide Research Data Management Policy. Educase Review. 12/6/2013. <http://er.educause.edu/articles/2013/12/starting-the-conversation-universitywide-research-data-management-policy>

## ***Next Steps***

With this report, the RDM working group officially completes its activities. Further steps from this point include:

- Continuing to maintain the RDM website. (Director of Digital Data Curation Strategies)
- As a follow-up to the RDM survey, conduct in-depth interviews with 12-15 investigators. (Director of Digital Data Curation Strategies)
- Implement recommendation for development of Best Practice guidance. (As directed by VP/IT).
- Convene a group with authority to establish priorities and a timeline as well as specific budget recommendations to support infrastructure and services development. (At the direction of the VP/IT and Senior Vice-Provost).

## ***Members of the RDM Working Group***

T. Scott Plutchak (co-chair)

Director of Digital Data Curation Strategies (Office of the Sr. Vice Provost)

David Yother (co-chair)

Asst. Vice President, Technology Services, UAB IT

Thomas Anthony

Director, Big Data Research and Analytics Laboratory

Lauretta Gerrity

Senior Associate VP - Research Administration, OVPRED

Jeff Graveline

Assistant to the Dean for Scholarly Communications and Copyright Compliance, UAB Libraries

Andrew Keitt

Associate Professor, Director of the Graduate Program, UAB Department of History

Kent T. Keyser

Asst. Vice President for Research, OVPRED

Elliott Lefkowitz

Director of Informatics, Center for Clinical and Translational Science

Tim Parker

Associate Director - Federal Area, Office of Sponsored Programs

Joseph Roberson

Associate University Compliance Officer

Lee Vucovich

Chair of Reference Services & Liaison to Graduate Biological Sciences, the CCTS, and the Joint Health Sciences Departments, UAB Libraries

## **Appendix A**

### **Supporting Compliance With Federal Funder Mandates for Public Access to Publications and Data Resulting from Federally Funded Research**

#### **Charge to the Working Group**

Following the release, in February 2013, of the OSTP memo “Expanding Public Access to the Results of Federally Funded Research,” the Federal agencies with over \$100 million in annual conduct of research and development are now making public their plans for complying with the OSTP directive. Details of the plans remain vague in many cases, but it is clear that over the course of the next 6 to 12 months, grant applications will need to include Data Management Plans (DMP) of varying levels of specificity.

If UAB investigators are to be competitive in acquiring grant funding and in publishing in top tier journals, it is essential that we develop a clear strategy for supporting investigators in responding to these requirements. Several units across campus have expertise in and responsibility for various aspects of that support. The task of the working group will be to develop recommendations for the specific responsibilities of, and coordination among, these units. The recommendations will go to Curt Carver, VPIT/CIO who will review and provide feedback and guidance to the working group for next steps.

The working group will be co-chaired by Scott Plutchak and David Yother, and will meet monthly for 4 to 6 months, with an initial meeting in early October.

The specific issues to be addressed are:

- Monitoring the development of the agencies’ plans – some agency’s plans are well developed. Most are not. NIH, in particular, is tasked with the challenge of overseeing many existing sets of requirements from different institutes and pulling together a comprehensive set of policies. We need to have someone tasked with keeping track of these policies as they change, assessing their impact on UAB support and workflows, and maintaining a website with updated information.
- Providing support for Data Management Plans. The DMPTool, developed by the California Digital Library, contains information and templates that can be helpful in writing acceptable data management plans. We can add UAB specific information to the Tool. As IT develops more infrastructure resources to support research we need to be sure that DMPs reflect appropriate use of UAB resources. We should have a mechanism for the DMPs to be reviewed during the submission process, but OSP does not currently have the expertise to do that. As IRAP development continues, we should investigate how DMP support can be incorporated into IRAP functionality.
- Develop a process for monitoring the implementation of DMPs during the life of the project. This should include assistance in developing appropriate



metadata and context for the data so that it can be appropriately deposited when necessary.

- At appropriate points during the project and upon its completion, data will need to be deposited in authorized repositories according to the requirements of the funders. We need to develop supportive guidance and workflows to assist investigators in doing this in a timely fashion.
- As articles are accepted for publication, public access versions will need to be handled according to the requirements of the funding agency (or agencies). We have a good model for that in the way that Lister Hill Library works with the Asst VP for Research to support the NIH Public Access policy. We will need to extend and expand that process to meet the requirements of other funders.
- Increasingly journals will require deposit of, or access to, the data underpinning the articles that they publish. We will need to develop appropriate guidance and support to investigators to manage those deposits as necessary. As is the case with funders, journal policies will vary and evolve.

September 2015

## **Appendix B**

### **Defining Research Data**

OMB circular A-110 defines *research data* “as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues.”

Research data comes in many forms, formats and sizes. While the basic principles of good data management apply to all, their application will vary. Several data management websites list the following as examples:

- Text or Word documents, spreadsheets
- Laboratory notebooks, field notebooks, diaries
- Questionnaires, transcripts, codebooks
- Audiotapes, videotapes
- Photographs, films
- Test responses
- Slides, artifacts, specimens, samples
- Collection of digital objects acquired and generated during the process of research
- Data files
- Database contents including video, audio, text, images
- Models, algorithms, scripts
- Contents of an application such as input, output, log files for analysis software, simulation software, schemas
- Methodologies and workflows
- Standard operating procedures and protocols

(See, for example, University of Oregon: Research Data Management; Boston University Libraries: Research Data Management)

## **Appendix C**

### **Recommendation for development of comprehensive guidelines for best practices in research data management.**

September 2016

The Research Data Management working group has identified as a critical need the development of comprehensive guidelines identifying best practices for the management of research data. In order to be effective, these guidelines need to be developed in a coordinated fashion involving appropriate representation from IT, OVPRED, University Compliance, the CCTS, and the Office of the Provost. We recommend that such a committee be charged as soon as possible with developing such guidelines and crafting a strategy for disseminating and encouraging their adoption throughout the institution.

The recent RDM survey of current data management practices confirms that the majority of the data that is generated by UAB researchers is maintained on departmental servers, individual laptops and flash drives, as well as in paper records. Individual labs and researchers are responsible for developing and maintaining their own management practices. At the institutional level we have few mechanisms for insuring that best practices for security, backups and preservation are being followed, or that proper training and guidance are being provided to new investigators and assistants.

The NIH's current emphasis on Rigor and Reproducibility, along with emerging federal funder policies requiring data management plans that emphasize data sharing, highlight the importance of improving our institutional approach to research data management and the need to provide adequate support to investigators in adhering to best practices.

Several relevant UAB documents currently exist:

- [Guidelines for Data Collection, Documentation, and Storage](#) (from OVPRED)
- [Data Custodian Responsibilities](#) (from IT)
- [Data Protection and Security Policy](#)
- [Interim Guidelines for the Use of Cloud Services on UAB Campus](#) (from IT)

These have been developed independently over a number of years to serve specific purposes and can serve as a good basis for developing a clear and comprehensive approach to guidelines that are easy to understand and straightforward to follow.

Once such consensus guidelines are developed it will be essential to develop a strategy for promulgating them widely and incorporating them into relevant training programs.

**Appendix D**  
**RDM Services at other universities**

While effective Research Data Management requires collaborative efforts among many units of the University, in most institutions it is the library that has taken the lead role in providing information and developing basic services.

UAB is currently ranked 6<sup>th</sup> in federal funding among SE universities. The table below provides a snapshot of the services provided at the institutions ranked above UAB. (Caveat: the descriptions are based on an examination of institutional websites. The extent of services provided and the degree to which they are used may vary considerably from institution to institution).

	UNC-CH	Duke	GA Tech	Vandy	Emory
Has some campus-wide policies related to RDM		X	X		
Provides information on funder requirements	X		X		X
Provides assistance with developing DMPs	X	X	X	X	X
Links to DMP Tool	X	X	X	X	X
Offers assistance in providing metadata for datasets	X		X		X
Provides information on RDM Best Practices	X	X	X		X
Provides training on RDM Best Practices	X	X	X		X
Central support for Research Data storage & security	X	X		X	X
Provides information on data repositories	X	X	X	X	X
Manages an institutional data repository	X	X	X		X
Provides assistance with data deposit	X				

Notes:

UNIV OF NORTH CAROLINA CHAPEL HILL (8)  
<http://guides.lib.unc.edu/researchdatatoolkit> Site maintained by UNC libraries; numerous campus partners including, among others, Office of the Vice Chancellor for Research, ITS Research Computing, Sheps Center for Health Services Research, Odum Institute for Research in Social Science. Provides full range of services including DMP assistance, training, long-term data storage.

DUKE UNIVERSITY (10)

<http://library.duke.edu/data/guides/data-management>

<https://rc.duke.edu/data-management/>

Duke Research Computing provides a variety of data management services. The libraries provide some support for addressing DMP requirements and data archiving. The libraries and Research Computing collaborate on data visualization services. Duke recently created 4 new positions within the library system to assist investigators with data management and curation.

GEORGIA TECH (11)

<http://d7.library.gatech.edu/research-data/home>

The libraries provide assistance with DMPs, training that satisfies RCR requirements, and help with finding appropriate data repositories (including SMARTech, the institutional repository.)

VANDERBILT UNIVERSITY (24)

<http://researchguides.library.vanderbilt.edu/datamanagement>

The library maintains a website with information on developing DMPs and data repositories. They have a data management consulting group that can be contacted for assistance.

EMORY UNIVERSITY (27)

<http://guides.main.library.emory.edu/datamgmt>

The library maintains a very detailed and comprehensive informational website and provides a broad range of data management services across the data life cycle. There appears to be good collaboration with other appropriate university units. Emory's Libraries & Information Technology Services Research Data Management Working Group issued a report in January 2015 with detailed recommendations for advancing RDM services locally ([http://guides.main.library.emory.edu/ld.php?content\\_id=8769380](http://guides.main.library.emory.edu/ld.php?content_id=8769380))

**Appendix E**  
**Research Data Management Survey**  
Final Report  
December 2016

Introduction

In order to establish a high-level perspective on the current state of Research Data Management at UAB, the RDM working group developed a brief survey that ran during May 2016. The survey, as expected, highlighted several areas where, because of the extreme decentralization of RDM management, UAB faces substantial risks in its management of research data, particularly with regard to compliance with federal mandates regarding data security and data sharing.

Some Key Findings:

- Investigators desire more data storage, particularly cloud storage that is usable by multiple people.
- There is a high need for expertise in database development, data management, and assistance with systems.
- Much research data is stored locally, on personal computers or departmental servers.
- There appears to be a lack of systematic backup protocols for research data.
- Estimates of costs of RDM vary widely – in many cases investigators are unable to calculate what is being spent.
- There is a lack of awareness of institutional resources available, leading to unnecessary duplication.

Methods

The survey was developed (using SurveyMonkey) during April 2016 in an iterative process involving members of the RDM working group and attendees at the monthly Data Wranglers discussion sessions. Using the Active Projects Detail Report in IRAP, a list of current PIs was generated. After deleting addresses for individuals no longer at UAB, there were 924 unique addresses. An email announcing the survey was sent to those addresses on May 5 and the survey link was sent on May 10. The survey was closed on May 24<sup>th</sup>. There were 118 responses for a response rate of 13%. While the response rate was too low to draw statistically valid conclusions the responses are certainly illustrative of the issues facing UAB investigators.

Results

The respondents received funding from a variety of government and non-government sources. Not surprisingly, 65% were funded by NIH.

The primary location reported for data storage was departmental servers (66%). Fifty-seven percent store data on the hard drive of a desktop computer and 44%

have data on paper. Only 14% use a UAB cloud service and 13% a non-UAB cloud service. Data management is performed by lab personnel in most cases (82%) with 36% using department or school IT personnel.

Funder required data-sharing is clearly not a major concern at this point. A third of respondents skipped the question about data sharing and in the comments, many of those who did respond indicated that it had not yet come up.

On the size of data being stored, 52% reported the total as between 1GB and 1TB and an additional 33% reported 1TB to 100TB. Only one respondent said they stored over 100TB of data.

Respondents were asked about the sensitivity of their data (low, medium, high). If their data fell into multiple categories they were asked to estimate the percentage of each. Fifty-six percent reported hosting data of medium sensitivity (defined as not public, confidential but not sensitive), while 45% have highly sensitive data (PII (SSN), PHI/HIPAA, FERPA, PCI).

The survey asked for an estimate of total annual expenses related to research data management (including hardware, software and personnel). Responses ranged from as little as \$0 to a high of \$1,000,000. Comments included “unknown” and “a lot.”

Respondents were asked an open-ended question: What services or other support could UAB provide to make it easier for you to manage your research data? Seventy-one respondents provided comments, frequently making multiple suggestions, for a total of 102 individual suggestions. A loose categorization of the comments shows 22 items asking for expert assistance with some aspect of data management and another 22 addressing the need for improved cloud storage solutions. The latter group is of particular interest. The survey was launched not long after UAB BOX was implemented, and it is clear from the comments that some respondents were not yet aware of it or did not fully understand its capabilities. The need for more inexpensive data storage (independent of comments specifying cloud services) was also noted by many as was the need for central back-up solutions. There were also comments addressing the need for training, better support for clinical data projects, improved data transfer speeds and other infrastructure related issues.

The final question simply asked if the respondents were willing to participate in a follow-up interview to assess their data management needs in more detail. Half of the respondents said yes. Follow-up interviews with 12-15 of these are planned for early in 2017.

## Conclusions

The results from the survey are useful in highlighting areas of concern among UAB investigators regarding the resources available to them in addressing their research data management needs. Data storage, expert assistance, better awareness of

resources available to investigators and the need for a better accounting of the costs of research data management are all indicated.

(Full survey results available on request)

T. Scott Plutchak  
Director of Digital Data Curation Strategies  
December 2016

**Acknowledgments:** Attendees at the monthly Data Wranglers sessions were instrumental in developing the final survey questions. Kathryn Kaiser, Ph.D. of the Office of Energetics, School of Public Health, provided substantial assistance throughout.