Data Needs Assessment

Research Data Management Group Survey of NSF Principal Investigators at Cornell University

Developed and administered by: Dianne Dietrich, Eric Chen, Florio Arguillas, Stefan Kramer, and Gail Steinhart

This document contains the survey instrument used in a 2011 survey of NSF PIs at Cornell University.

Block 0

As you may be aware, on October 1, 2010, the National Science Foundation announced a new policy requiring a supplementary document for all grant proposals outlining the proposal's data management plan (see http://www.nsf.gov/bfa/dias/policy/dmp.jsp). This requirement will take effect on January 18, 2011. Individual programs and directorates within NSF may have additional guidelines. Other major research funders can be
expected to implement similar policies, if they have not already done so.

The Research Data Management Service Group (RDMSG, http://data.research.cornell.edu/) is conducting this survey to estimate the demand on campus services for data management, and to identify potential gaps in existing services.

It should take you approximately 10 minutes to complete this survey and your participation is voluntary. You will not be required to provide any identifying information unless you choose to.

Your answers will provide valuable information for use in the RDMSG's planning efforts. Some results from this survey, such as general trends, may be used in external reports, but no identifying information or direct quotes will be used without your consent.

This survey will be closed and no further submissions will be accepted after February 1st, 2011.

Information Sessions

You are also invited to attend an informational session on the National Science Foundation's (NSF) new policy requiring a data management plan with all grant proposals. The new policy goes into effect January 18, 2011.
Staff from RDMSG will review the new requirement, describe how researchers can obtain assistance from the RDMSG to create data management plans, and answer questions.

Three sessions will be offered:

Thursday, January 13, 1:30-2:30pm, G01 Biotech
Tuesday, January 18, 9:00-10:00am, 102 Mann Library
Thursday, January 20, 12:30-1:30pm, 312 Hollister

By clicking the next button below, you voluntarily agree to participate in this online survey.

Block 1

Please answer the following questions with your most recent NSF award in mind.

Please specify the NSF directorate of your most recent award.

- Directorate for Biological Sciences
- Directorate for Computer & Information Science & Engineering
- Directorate for Education & Human Resources
- Directorate for Engineering
- Directorate for Geosciences
- Directorate for Mathematical & Physical Sciences
- Directorate for Social, Behavioral & Economic Sciences
- Office of the Director (includes Office of Cyberinfrastructure, Polar Programs, and others)

Would you be interested in any sort of guidance, including consultation, for writing a data management plan in support of an NSF grant application?

- Yes
- No
- I'm not sure

Additional comments

Block 2

According to the NSF, a data management plan may include a description of "the types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project."

Please specify the types of data you have produced or anticipate producing for this project that you intend to share with others. Check all that apply.

- [ ] Text
- [ ] Image
- [ ] Audio
- [ ] Video
- [ ] Spreadsheets
- [ ] Databases
- [ ] Code
- [ ] Other
- [ ] I'm not sure

Please specify other data types

Please list the file extensions you have produced or anticipate producing for this project that you intend to share with others.

Block 3

According to the NSF, a data management plan may include a description of "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)."

Does the data you have produced or intend to produce conform to known standards in your discipline?

- Yes
- No
- I'm not sure

Please specify the standard(s) you are using.

"Metadata" refers to descriptive information or documentation about data.

Have you produced or do you anticipate producing metadata for this project?

- Yes
- No
- I'm not sure

Additional comments

Does the metadata you have produced or intend to produce conform to known standards in your discipline?

- Yes
- No
- I'm not sure

Please specify the standard(s) you are using.

Would you make use of a service to produce metadata for this project?

- Yes, and I would be willing to pay for this service
- Yes, but I would not be willing to pay for this service
- No, I would produce metadata myself

Block 4

According to the NSF, a data management plan may include a description of "policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements."

Do you anticipate needing to consult with an intellectual property specialist to create a license agreement or usage statement for the data you have produced or intend to produce?

- Yes
- No
- I'm not sure

When would you be able to share the data you have produced or intend to produce for this project?

- Immediately after collection
- Immediately after my team has analyzed the data
- Six months or more after my team has analyzed the data
- I would not be able to share this data

What might prevent you from sharing the data you have produced or intend to produce for this project? Check all that apply. (You may also check no boxes if none apply.)

- [ ] Little value to others
- [ ] Confidentiality or privacy issues
- [ ] Commercialization or patent issues
- [ ] Some or all of the data I work with has license or usage restrictions that prevent me from sharing
- [ ] Data requires secure access I am not capable of providing

Additional comments

Block 5

According to the NSF, a data management plan may include a description of "policies and provisions for re-use, re-distribution, and the production of derivatives." Furthermore, Investigators 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."

Given the NSF expectation to share data with other researchers, how much data would you intend to share?

- I do not plan on sharing data
- No more than 1 GB
- More than 1GB but less than 100 GB

- More than 100 GB but less than 1 TB
- More than 1 TB but less than 100 TB
- More than 100 TB

When you publish your findings from this research project, do you plan on submitting your supporting data to a journal publisher?

- Yes
- No
- I'm not sure

Additional comments

Do you plan on using a custom solution to share the data you have produced or intend to produce? (i.e., Sharing data on a personal or departmental website or FTP server…)

- Yes, and I plan to do this work in-house
- Yes, and I plan to contract all or part of this work
- No
- I'm not sure

Additional comments

Block 6

According to the NSF, a data management plan may include a description of "plans for archiving data, samples, and other research products, and for preservation of access to them."

Do you plan to deposit the data you have produced or intend to produce in Cornell's Institutional Repository, eCommons (http://ecommons.cornell.edu/about.html), or would you be interested in doing so to satisfy the NSF requirement?

- Yes
- No
- I'm not sure

Additional comments

Do you plan to deposit the data you have produced or intend to produce in CISER's Data Archive (http://ciser.cornell.edu/info/about.shtml), or would you be interested in doing so to satisfy the NSF requirement?

- Yes
- No
- I'm not sure

Additional comments

Do you plan to utilize the Cornell Restricted Access Data Center http://ciser.cornell.edu/CRADC /What_is_CRADC.shtml) to work with restricted access or limited use licensed data, or would you be interested in doing so to satisfy the NSF requirement?

- Yes
- No
- I'm not sure

Additional comments

Do you plan to store the data you have produced in the Center for Advanced Computing Disk Farm (http://www.cac.cornell.edu/services/storage.aspx), or would you be interested in doing so to satisfy the NSF requirement?

- Yes
- No
- I'm not sure

Additional comments

Do you plan to deposit the data you have produced or intend to produce in a data center or other non-Cornell repository, or would you be interested in doing so to satisfy the NSF requirement?

- Yes
- No
- I'm not sure

Please specify the repository (or repositories) you plan to deposit your data into.

Block 7

What is your current method of backing up the data you have produced or intend to produce for this project? Check all that apply.

- Own IT infrastructure (e.g., external hard drives)
- EZBackup or other campus-based solution
- Commercial solution (i.e., Google Docs, Amazon S3
- No backup

Approximately how much data needs to be backed up?

- No more than 1 GB
- More than 1 GB but less than 100 GB
- More than 100 GB but less than 1 TB
- More than 1 TB but less than 100 TB
- More than 100 TB

Block 8

The NSF specifies that if "any PI or co-PI identified on the project has received NSF funding in the past five years, information on the award(s) is required." Specifically, applicants must indicate "evidence of research products and their availability, including, but not limited to: data, publications, samples, physical collections, software, and models, as described in any Data Management Plan."

Do you currently keep track of research outputs and their availability?

- Yes
- No

Additional comments

If there was a service offered where you could enter in basic information about your data (including the description, where it was available on the web) to demonstrate compliance with NSF's policy, would you make use of it?

- Yes
- No
- I'm not sure

Additional comments

Do you anticipate or would you be interested in any sort of guidance, including consultation or instruction, for any of the data management plan components mentioned above?

- Yes
- No
- I'm not sure

Additional comments

Which components are you interested in receiving consultation or instruction for?

A review of the datamangement components:

1. the types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;
2. 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);
3. policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
4. policies and provisions for re-use, re-distribution, and the production of derivatives; and
5. plans for archiving data, samples, and other research products, and for preservation of access to them.

Block 9

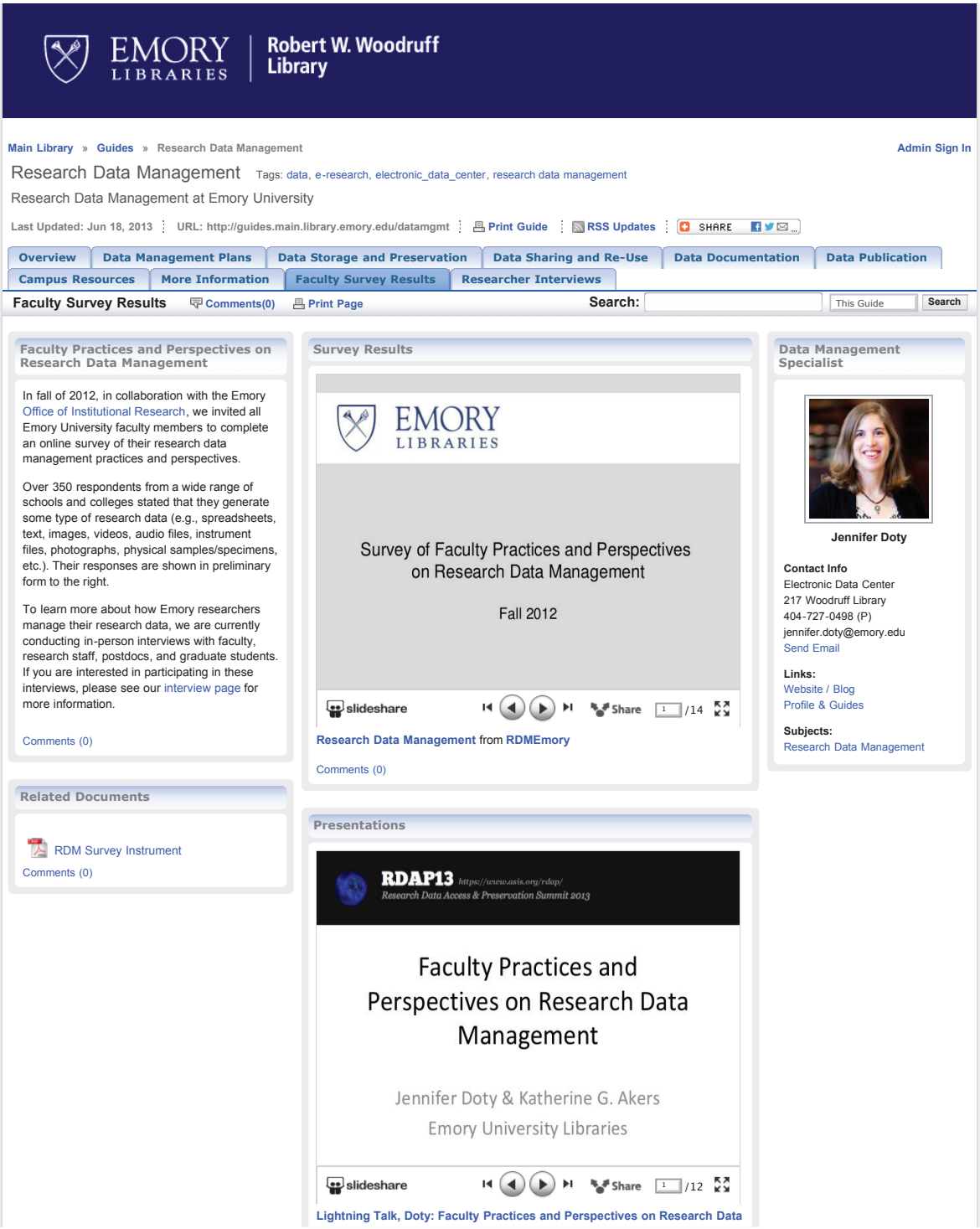
A review of the datamangement requirements:

1. the types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;
2. 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);
3. policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
4. policies and provisions for re-use, re-distribution, and the production of derivatives; and
5. plans for archiving data, samples, and other research products, and for preservation of access to them.

Please share any additional thoughts or concerns you have regarding campus support for complying with funders' data management policies.

Would you like to be contacted in the future about your response or participate in focus groups about Data Management Plans? (Please include your contact information if yes.)

EMORY UNIVERSITY
Research Data Management. Faculty Survey Results
http://guides.main.library.emory.edu/datamgmt/survey

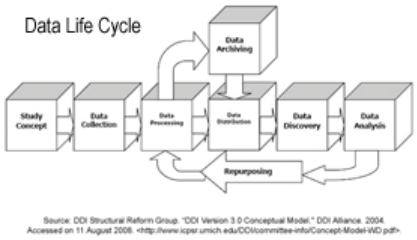


Data Management and Publishing

Home
Why Manage Your Data?
Data Planning Checklist
What is Data?
Evaluate Your Data Needs
Funding Requirements
Data Management Plans
Writing an NSF Data Management Plan
Documentation and Metadata
File Formats
Organizing Your Files
Backups and Security
Sharing Your Data
Citing Data
Data Integration
Ethical and Legal Issues
Workshops
Guides to Data Management
Managing Research Data 101 (PDF)
Inquiries?

Related Guides:

A Data Planning Checklist



Source: DDI Structural Reform Group. "DDI Version 3.0 Conceptual Model." DDI Alliance. 2004. Accessed on 11 August 2008. <http://www.icpsr.umich.edu/DDI/committee-info/Concept-Model-WD.pdf>

Managing your data before you begin your research and throughout its life cycle is essential to ensure its current usability and long-run preservation and access. To do so, begin with a planning process. See also our page on data management plans.

1. What type of data will be produced? Will it be reproducible? What would happen if it got lost or became unusable later?
2. How much data will it be, and at what growth rate? How often will it change?
3. Who will use it now, and later?
4. Who controls it (PI, student, lab, MIT, funder)?
5. How long should it be retained? e.g. 3-5 years, 10-20 years, permanently
6. Are there tools or software needed to create/process/visualize the data?
7. Any special privacy or security requirements? e.g., personal data, high-security data
8. Any sharing requirements? e.g., funder data sharing policy
9. Any other funder requirements? e.g., data management plan in proposal
10. Is there good project and data documentation?
11. What directory and file naming convention will be used?
12. What project and data identifiers will be assigned?
13. What file formats? Are they long-lived?
14. Storage and backup strategy?
15. When will I publish it and where?
16. Is there an ontology or other community standard for data sharing/integration?
17. Who in the research group will be responsible for data management?

Faculty Successes:

"I've had thousands of downloads of my published data--I am impressed that it's been so useful to others!"

Esther Duflo, Abdul Latif Jameel Professor of Poverty Alleviation and Development Economics, MIT

For advice on a data management project, contact:

data-management@mit.edu

Courtney Crummett
Bioinformatics and Biosciences Librarian

Anne Graham
Civil and Environmental Engineering, Building Technology Librarian

Katherine McNeill
Social Science Data Services & Economics Librarian

Daniel Sheehan
Senior GIS Specialist

Amy Stout
Electrical Engineering and Computer Science Librarian

PURDUE UNIVERSITY
Data Curation Profiles Toolkit
http://datacurationprofiles.org/

Data Curation Profiles Toolkit

About | News | Submit a Profile | Completed Profiles | Workshops | Forums | Resources

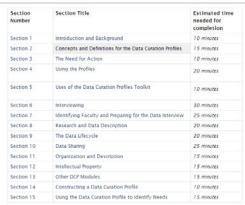
Welcome to the Data Curation Profiles Community!

A LOT IS GOING ON WITH DATA CURATION PROFILES: THREE (3) NEW TOOLS!

We are in the middle of renovating this space, but we felt we had to share!

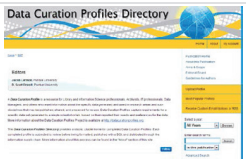
The Data Curation Profiles Symposium was recorded and provides a video overview of work involving the Profiles and Toolkit. Additional presentations by experts in the field addressing curation.

http://docs.lib.purdue.edu/dcpsymposium/



A new tutorial on using the Data Curation Profiles is available to anyone who wants to learn more about the Profiles and the Toolkit.

Coming soon…



We have a new publication, the Data Curation Profiles Toolkit. You can publish Profiles you write, they can be found and studied, and they will be indexed to be easily found and cited.

http://docs.lib.purdue.edu/dcp

LOGIN

Download the Toolkit

PURDUE UNIVERSITY LIBRARIES

INSTITUTE of Museum and Library SERVICES

D2C2 DISTRIBUTED DATA CURATION CENTER

This website is an environment where academic librarians of all kinds, special librarians at research facilities, archivists involved in the preservation of digital data and those who support digital repositories can find help, support and camaraderie in exploring avenues to learn more about working with research data and the use of the Data Curation Profiles Toolkit.

A Data Curation Profile is essentially an outline of the "story" of a data set or collection, describing its origin and lifecycle within a research project. The Profile and its associated Toolkit grew out of an inquiry into the changing environment of scholarly communication, especially the possibility of researchers providing access to data much further upstream than previously imagined. If researchers are interested in sharing or forced to provide access to data sets or collections, what does that mean for the data, for researchers, and for librarians?

Data Curation Profiles can:

- provide a guide for discussing data with researchers
- give insight into areas of attention in data management
- help assess information needs related to data collections
- give insight into differences between data in various disciplines
- help identify possible data services
- create a starting point for curating a data set for archiving and preservation

Look around and get to know the site. You will find everything from the history of Data Curation Profiles, the Toolkit for developing a profile of a research data set (registration is required), completed profiles from various disciplines, guidelines for submitting profiles, forums for discussion and resources to learn more about data curation. We hope you will register, download the Toolkit, submit a Profile of your own and join the conversation.

Data Interview Protocol

This document is the step-by-step set of instructions for the actual interview. This is to serve as the master copy, accompanied by a question template that is designed to be printed and used to ask focused questions along with check boxes to account for all of the protocol issues. The template will also allow for note taking during the interview.

Data Interview Constraints

- Interview will consist of:
  - Scientific Data Consultant Group members (two) and subject librarian (one)
  - Researcher (one) being interviewed.
  - Optionally, an additional technical expert invited by the researcher.
- An interview will last no more than sixty minutes.
- Interviews will be semi-structured to allow free-flowing discussion.
- Information to be gathered includes:
  - The state of current data management efforts.
  - Types of digital data created.
  - A prioritized needs assessment covering:
    - Current situation and future needs.
  - Functional specifications for services to meet those needs.

Mission of the Data Interview

At the start of the interview, we will briefly review why we are doing these data interviews and why the library is suited to do it.

- Library goal of supporting researcher needs.
- Library focus on data management.
- Scientific Data Consultant Group Experience:
  - Research Computing Lab, Dataset Task Force, Metadata Steering Group, Institutional Repository Implementation Team
- Purpose of the Data Interview Initiative:
  - Identify common researcher data problems and needs.
  - Identify communities and individuals who are under the most pressure from upcoming grant regulations.
  - Provide data management recommendations and training.
  - Identify the types of digital "data" that are being created.
  - Identify potential partnerships for IR data deposit implementation.
- Remember – there are no "right" answers! We want an honest assessment of your practices. That includes your successes and your failures.
- Mention IRB and give them a copy

Copyright (c) 2012 by UVa Library. This work is made available under the terms of the Creative Commons Attribution-ShareAlike 3.0 license, http://creativecommons.org/licenses/by-sa/3.0/.
© 2012 by the Rector and Visitors of the University of Virginia

Data Interview Protocol

What Is Your Data All About?

To start the interview we'd like to get some background information on your research. If you'd like, you can discuss your lab's work as a whole, or focus on a specific project.

1.1 What question are you trying to answer?
1.2 What is the process/method to answer the questions?

What Kind of Data Do You Have?

Now that we've heard about your research, let's talk specifically about what kind of data you produce.i.e. what they create and use, and their attitude towards digital material. Here we are looking for the data characteristics, types, sizes and transformations.

2.1 Describe the data you create in your research.
Here we are looking for the data characteristics, types, sizes and transformations.

- ❑ General Category (experimental, simulation/computational, observational, derived/compiled)
- ❑ Creation (sensors, instruments, software)
- ❑ Data Type (docs, emails, databases, images, videos, etc.)
- ❑ Data Format (MS Word, Excel, spss, html, jpg, etc.)
- ❑ Amount (#files, files sizes, growing?)

2.2 Another issue related to data is that of intellectual property. Who owns the Intellectual Property rights of the data you create? Are you familiar with the following UVa policies?

- ❑ Lab Notebook Policy
- ❑ UVa's Ownership Rights Policy

How Do You Work With Your Data?

Now we'd like to talk about the practices you have in place to organize your data.

3.1 Who is responsible for managing the data? Are you using any filing or naming conventions for the files? How are the files organized? Is there any documentation on the files and/or data fields?

Copyright (c) 2012 by UVa Library. This work is made available under the terms of the Creative Commons Attribution-ShareAlike 3.0 license, http://creativecommons.org/licenses/by-sa/3.0/.
© 2012 by the Rector and Visitors of the University of Virginia

Data Interview Protocol

Here we are looking for information on managing the data. Are there set procedures? What role does each person play?

- ❑ Management Plan
- ❑ Naming Conventions
- ❑ File Organization
- ❑ Documentation
- ❑ File Backup/loss/recovery
- ❑ File storage
- ❑ Backups

3.2 Do you share data among lab group or other colleagues (e-mail, shared drive, removable devices, CD, web pages, other)? Do you typically have multiple people working on the same data files? If so, have you had issues regarding which version was "correct" or the latest? How are these issues controlled or resolved?

- ❑ File sharing
- ❑ Issues related to multiple file versions

Preservation Concerns

We are looking for any digital preservation issues in this section of the interview. Continue discussion to ascertain whether any issues have been encountered when creating and using digital material to identify areas where practices could improve.

Here we are looking for preservation issues on their own data in their own lab/computer.

4.1 What challenges have you faced in terms of storage, formats, costs, and continued access to older data?

- ❑ Do they have older files?
- ❑ Obsolete data formats
- ❑ Obsolete media
- ❑ Lost or misplaced data
- ❑ Storage space
- ❑ Costs

Copyright (c) 2012 by UVa Library. This work is made available under the terms of the Creative Commons Attribution-ShareAlike 3.0 license, http://creativecommons.org/licenses/by-sa/3.0/.
© 2012 by the Rector and Visitors of the University of Virginia

Data Interview Protocol

Data Sharing and Long-term Accessibility

Get them thinking about the future of their data i.e. how can these files continue to be accessed and used (if appropriate), do they need to be preserved, if so, for how long?

5.1 Have you been asked to provide or share your data? Could or should your data be reused or repurposed by others, and if so, how and by whom?

- ❑ Publisher requirement
- ❑ Funder requirement
- ❑ Restrictions (Confidentiality, Sensitivity)
- ❑ Documented for sharing

Long-term Preservation

5.2 Do your files need to be preserved? For how long? Does all of it need to be kept?

- ❑ Raw or processed data or both
- ❑ Who decides? Who is responsible?
- ❑ Where?
- ❑ Libra, the UVa IR
- ❑ How long?

What Would Make Data Management Easier for You?

Ask where the interviewee currently gets advice and support and what else s/he would like to see provided by the University. Key thing is to gauge desire for preservation policy, suggested coverage and any supplementary support needed to implement it.

6.1 What would help you create and manage your data better?

6.2 Who should be responsible for digital preservation? Who should be responsible for funding it?

- ❑ Preservation responsibility
- ❑ Help, where?
- ❑ Library

6.3 What sort of impact might a University-wide policy on data preservation have upon you? What sort of policy do you think would be reasonable?

Copyright (c) 2012 by UVa Library. This work is made available under the terms of the Creative Commons Attribution-ShareAlike 3.0 license, http://creativecommons.org/licenses/by-sa/3.0/.
© 2012 by the Rector and Visitors of the University of Virginia

Data Interview Protocol

Follow-Up Plans

Review the steps that will come after the interview is complete (Script for interview is included below).

7.1 Team combines interview notes.
7.2 Send aggregated report to researcher for review/approval, corrections/additions on notes. To expediate things we need the approval/feedback back within one week.
7.3 Ask for feedback for interview process.
7.4 Provide a complete report that includes a summary of the conversation, responses to the interview questions, and recommendations on how to improve your data management.

Script from the Data Interview Template (for this section):

Thank you for participating in our Data Interview. Here are our next steps:

7.1 Andrew, Sherry and I will combine our interview notes.
7.2 I will send you an aggregated report for your review/approval, corrections/additions on our notes. Please return the approval/feedback within one week.
7.3 When you send the report back to us, we would like to have your feedback on our interview process.
7.4 Once we have your comments on the report. I will provide you with a complete report that will include a summary of the conversation, responses to the interview questions, and recommendations on how to improve your data management.

Copyright (c) 2012 by UVa Library. This work is made available under the terms of the Creative Commons Attribution-ShareAlike 3.0 license, http://creativecommons.org/licenses/by-sa/3.0/.
© 2012 by the Rector and Visitors of the University of Virginia