


Research Data Management DataGuide

Version 4.1

Contact – Eugene Barsky, Research Data Librarian, UBC Library,
eugene.barsky@ubc.ca

Last Updated - 12 July 2016

Introduction

Take a moment to think about your research and the kinds of data you generate.

- Where is this data stored and how is it organized?
- If you were asked to share your data with another researcher would they be able to make sense of your work?
- If you needed to locate your data files from 5 years ago, how easy would they be to find and use?

If you are unsure about the answer to any of these questions you have come to the right place. The following series of modules was created for graduate students, research assistants and other researchers interested in improving their data management skills. The modules cover the basics of data management plans, metadata and data documentation, data storage and security, and data sharing and are designed to get researchers thinking about the steps they can take to better manage their research data.

Module 1: Research Data and Data Management Plans

What is research data?

Before we begin our discussion of research data management we should clarify what we mean by research data. Research data can be defined as "a collection of facts, measurements, or observations used to make inferences about the world we live in" (Eisner and Vasgird, n.d.). In practice, research data can be a great many things, from DNA samples to interview transcripts to photographs.

Why is it important to properly manage your data?

Most researchers today work with a lot of research data, and without proper data management it can be difficult to keep track of everything!

Proper data management can make it easier for you to:

- Find your files
- Keep track of different versions of your data
- Organize and compile information at the end of a project
- Reproduce your work (if required for a journal or patent)
- Pass on your work to another researcher
- Share your work
- Satisfy grant requirements
- Satisfy journal requirements
- Satisfy research ethics board requirements

The research data lifecycle

It is important to remember that managing your research data is not something that you do at the end of a project, but throughout each stage of the process. The following diagram illustrates the different stages in working with research data:

http://www.data-archive.ac.uk/create-manage/life-cycle

In order to determine how best to manage data through each stage of a project, many researchers create a data management plan.

What is a data management plan?

A data management plan is a document that helps you manage your research data by planning out what you will do with your data during and after your research (DataOne, 2012).

A detailed data management plan can help you save time and money by getting you to think about the different steps in your research process and the resources and tools you will need to organize, store and share your data now and in the future.

Canadian requirements

Data management plans are listed as a key element in the June 2016 policy document - **Tri-Agency Statement of Principles on Digital Data Management** - http://www.science.gc.ca/default.asp?lang=En&n=83F7624E-1

Many American funding agencies including the National Science Foundation already require data management plans, so it is likely only a matter of time before they are a requirement for Canadian federal funds. We recommend that you save yourself a headache later and work on your data management plan now.

Components of a data management plan

A typical data management plan includes information about:

- Types of data produced
- Metadata or documentation standards
- Data security and encryption
- Data storage
- Intellectual property rights
- Data sharing
- Data archiving

There are many great resources for preparing a data management plan. The following checklists can help you decide what to include in your plan:

UK Data Archive Checklist: http://data-archive.ac.uk/create-manage/planning-for-sharing/data-management-checklist

These online tools walk you through the process of creating a plan:

- In Canada - **DMP Assistant**: https://assistant.portagenetwork.ca/
  Hosted by the collaborative Canadian Portage initiative, DMP Assistant is designed to meet the Data Management Plan recommendations (in English or French) of these Canadian funders:
  - Canada Foundation for Innovation (CFI)
  - Canadian Institutes of Health Research (CIHR)
  - Natural Sciences and Engineering Research Council (NSERC)
  - Social Sciences and Humanities Research Council (SSHRC)

- In USA - Data Management Plan Tool: https://dmptool.org

Module 1 References

DataOne (2012). *DataOne Education Module: Why Data Management?* Retrieved from http://www.dataone.org/sites/all/documents/L01_DataManagement.pptx

Digital Curation Centre (n.d.) *Data Management Plans.* Retrieved from http://www.dcc.ac.uk/resources/data-management-plans

Eisner, R., & Vasgird, D. (n.d.) Foundation Text. In *RCR Data Acquisition and Management.* Retrieved from http://ori.hhs.gov/education/products/columbia_wbt/rcr_data/foundation/index.html

Krier, L., & Strasser, C. A. (2014). *Data management for libraries: A LITA guide*. Chicago: ALA TechSource.

UK Data Archive (n.d.) *Create and Manage Data: Planning for Sharing.* Retrieved from http://www.data-archive.ac.uk/create-manage/planning-for-sharing

University of Minnesota Libraries (n.d.) *Data Planning Checklist.* Retrieved from https://netfiles.umn.edu/ul/Divisions/AcaProg/PhysSciEng/PSEShared/Instruction/Data%20Management/Public/Data%20Management%20Checklist%20Workshop.pdf

Module 2: Metadata and Data Organization

What is metadata?

Metadata is often described as "data about data" and helps answer the questions who, what, when, where, why.

Metadata makes it easier for researchers to:

- share their data
- publicize their data
- locate and retrieve data sets from others

Metadata can generally be divided into the following three categories:

Descriptive: Descriptive metadata describes the content and context of your data at both the dataset and item level.

Examples: title, author, keywords

Administrative: Administrative metadata includes information needed to use the data.

Examples: software requirements, copyright

Structural: Structural metadata describes how different data sets relate to one another, or any processing or formatting steps that were undertaken.

Examples: Information about the relationship between data sets in a database, file formats

Take a moment to think about your research project. What kind of descriptive, administrative and structural metadata might you want to record?

What are metadata standards?

Odds are you have already written down a good deal of metadata about your project; hopefully you don't plan on doing it all at the end. Save yourself some trouble and start gathering metadata at the beginning of your project.

Confused about what to record? Many disciplines have created their own metadata standards to ensure that data records can be interpreted and compared across projects and fields. A typical metadata standard provides a set structure and language for describing your data. Some of the most common metadata standards include Dublin Core, Darwin Core (for the biological sciences), and DDI (data documentation initiative).

If you are deciding which metadata standard to use remember that many data repositories, organizations, and journals have specific requirements for metadata. Double check before you commit.

Curious? Take a moment to visit the following link and find a metadata standard used in your field: http://www.dcc.ac.uk/resources/metadata-standards

Best practices in data documentation

Even if you decide not to follow a metadata standard it is still important to properly document your data. At the very least you should document the following in a readme.txt file stored alongside your data:

- context of data collection (the goal of your research)
- data collection methods (protocols, sampling, instruments, coverage…)
- structure of files
- sources used
- quality assurance (data validation, checking)
- data modifications
- confidentiality and permissions
- names of labels and variables
- explanations of codes and classifications

In addition to what you write, how you write it is very important. Always remember to be as clear as possible! The following are a list of best practices related to data documentation:

- Don't use jargon
- Define terms and acronyms
- State limitations

- Use descriptive titles
- Be specific and quantify
- Use keywords
- Make it machine readable (avoid symbols)

Finally, don't wait to document your data! If you wait until the end of your project you might lose valuable information!

File naming

Downloadable instructions – http://researchdata.library.ubc.ca/organize/

Large research projects can generate hundreds of data files. Proper file names and organization can make these files easier to locate and navigate.

It is recommended that you choose a file naming convention and stick with it. Make sure that everyone on your team is following the same rules for naming files. When deciding how to name your files remember the following:

- Keep file names under 32 characters
- Classify broad types of files (transcript, photo, etc.)
- Avoid spaces and special characters
- Use underscores instead of periods or spaces
- Make sure that file names are descriptive outside of their folders (in case they are misplaced or change locations); i.e., the file name should include all necessary descriptive information
- Include dates and format them consistently (international standard for date notation is YYYY_MM_DD or YYYYMMDD)
- Include a version number to track multiple versions of a document
- Be consistent!

Module 2 References

DataOne (2012). *DataOne Education Module: Metadata.* Retrieved from http://www.dataone.org/education-modules

DMP Tool (2014) *Data Management General Guidance*. Retrieved from https://dmptool.org/dm_guidance

Krier, L., & Strasser, C. A. (2014). *Data management for libraries: A LITA guide*. Chicago: ALA TechSource.

North Carolina Department of Cultural Resources (2008) *Best Practices for File Naming*. Retrieved from http://www.ncdcr.gov/Portals/26/PDF/guidelines/filenaming.pdf

UK Data Archive (n.d.) *Create and Manage Data: Documenting Your Data*. Retrieved from http://www.data-archive.ac.uk/create-manage/document

University of Oregon Libraries (n.d.) *Metadata and Data Documentation*. Retrieved from http://library.uoregon.edu/datamanagement/metadata.html

Module 3: Data Storage and Security

Data storage

Data storage and security considerations are essential aspects of managing research data and should be mapped out in your data management plan.

At the beginning of any project researchers should map out what data they will be generating and how they plan on storing it. In deciding where to store your data ensure that you understand your organization's policies and infrastructure for data storage and backups. A best practice is to have three copies stored in at least two locations (in case of a failure at one location).

Remember, just because you have saved your data doesn't mean it is safe! Data can be lost for a number of reasons including:

- hardware failures
- software failures
- viruses or hacking
- power failures
- natural disasters
- human error

Even if you are backing up your data, remember to check that the backups are working and that the data is accessible.

Finally, remember to attach your metadata alongside your stored data in order to ensure that everything can be understood in the future.

File types

Downloadable instructions – http://researchdata.library.ubc.ca/format/

File types are an essential consideration in data storage. Software and data storage technology changes quickly, and files can easily become obsolete or difficult to access. In general it is recommended that data files are copied to new media if technology changes or if files begin to degrade (every 2-5 years).

When choosing how to save your files chose a format that is:

- Non-proprietary
- Open, with documented standards
- Used by your community
- Encoded using standard character encoding
- Uncompressed

Understandably, many researchers prefer to work with proprietary file types like .doc or .xls. As these are not ideal for long-term data storage it is recommended that researchers save a version using open data formats for archiving.

The UK data archive has a chart of recommended file formats for a variety of data types: http://www.data-archive.ac.uk/create-manage/format/formats-table

Data security

As an ethical researcher, data security is an essential aspect of data management. Security regulations will differ based on the confidentiality of your data. Generally the more confidential your data, the more you should limit access to it.

Security planning should encompass the following areas:

- Network security: who has access to the network? Are there firewalls?
- Physical security: who has access to the computers? Who can access physical files? How is data transported?
- Computer security: Is anti-virus software up to date? Are you protected against power surges? Do you use passwords and firewalls? Is data encrypted? Is data storage secure?

If you are dealing with private or sensitive data make sure you understand your organization's regulations about storage, security, and disposal. Some countries including Canada do not allow personal data to be stored in servers outside the country, making commercial storage systems like Dropbox unusable.

Finally, remember that just because you deleted something doesn't mean it can't be recovered! To destroy data you must overwrite a hard drive, physically destroy memory sticks and shred paper documents.

Module 3 References

DMP Tool (2014) *Data Management General Guidance.* Retrieved from https://dmptool.org/dm_guidance

Krier, L., & Strasser, C. A. (2014). *Data management for libraries: A LITA guide*. Chicago: ALA TechSource.

UK Data Archive (n.d.) *Create and Manage Data: Storing Your Data.* Retrieved from http://www.data-archive.ac.uk/create-manage/storage

University of Oregon Libraries (n.d.) *File Formats*. Retrieved from http://library.uoregon.edu/datamanagement/fileformats.html

Module 4: Data Sharing and Re-Use

Why share data?

Have you considered what you might do with your data once your project has finished? Ever thought that someone else might benefit from your raw data? You might want to consider sharing your data!

Sharing research data can:

- Satisfy journal requirements
- Satisfy grant requirements
- Make research more open
- Promote scholarly rigor
- Raise the profile of a researcher
- Increase research efficiency
- Promote collaboration
- Establish a public record
- Maximize transparency
- Promote inquiry and innovation
- Increase the economic and social impact of research
- Provide greater resources for education and training

If that weren't incentive enough the Canadian Social Sciences and Humanities Research Council (SSHRC) and Canadian Institutes of Health Research (CIHR) now require grantees to deposit their data in publically accessible repositories.

Challenges to sharing data

While sharing research data can have huge benefits there are sometimes barriers to sharing. Preparing data for a repository can be time consuming and concerns about legal and ethical issues can make researchers wary of sharing data with others. Some types of data are simply not meant to be shared. These include trade secrets, medical information, commercial information, preliminary analysis, third party data, and geospatially linked data. Other data, however, can be shared after it has been anonymized.

In order to ensure you are sharing data in an ethical manner you should:

- evaluate the anonymity of your data
- obtain a confidential review (someone from the repository looks it over)
- comply with institutional regulations (e.g., those of your institution's research ethics board)
- comply with other regulations (HIPAA, BREB)
- have informed consent for data sharing
- restrict use of confidential data

How to share data

Once you have decided you are interested in sharing your data, how do you go about sharing it?

The easiest ways to share your data are to:

- Submit it into a subject specific repository or archive
- Submit it to an institutional archive
- Post it to a project website
- Submit it to a journal
- License your data and provide suggested data citation

Data repositories are an especially great way to share data as many of them offer long-term storage and preservation, regular backups, licensing arrangements, and online discovery and data promotion.

Data repositories exist at the institutional, national, and discipline level. It's probably a good idea to check with your colleagues and peers to see whether there is a recommended repository in your field.

When choosing a repository, consider the following:

- Who might want access to your data and where will they look?
- Is there an appropriate discipline specific repository?
- What are the access policies?
- What is the storage and preservation plan?

- What kind of data do they accept?
- What metadata standards are required?
- Do they charge any fees?

Take a moment to check out the repository database at http://www.re3data.org/. Is there a repository in your area? What are the requirements for submitting data?

Licensing data

If you decide to submit your data to a repository it is a good idea to license your data. Licensing data allows researchers to clearly state how they want their data to be used and makes it easier for others to re-use the data. While data itself does not fall under copyright protection, datasets and databases do, and the easiest way to protect your copyright while allowing access is by attaching a license.

Before deciding what license to use, you must first ensure that you yourself have permission to license the data, as only the rights holder can grant a license. Once you are sure you can grant a license, you must choose which license to apply. Make sure to check with your organization or repository as they might recommend a certain license or provide one for you.

The most common data licenses are from Creative Commons and the Open Data Commons. Both of these organizations have standard sets of licenses that allow data to be used in different ways.

Alternatively you can apply a waiver to place your data in the public domain and allow free and unrestricted access. The Creative Commons zero license is the most popular copyright waiver.

If you are interested in learning more about data licenses the Digital Curation Centre has an excellent guide to licensing research data:
http://www.dcc.ac.uk/resources/how-guides/license-research-data

Citing data

Let's say you are interested in using someone else's data that you have located in a repository. How do you cite it? While there is no official format for citing data, many journals and conferences have established data citation rules. Generally, it is a good idea to include the following information:

- Author/creator
- Date created
- Title
- Publisher
- Identifier (e.g. DOI or handle)

Module 4 References

Canadian Institutes of Health Research. (2013). *CIHR Open Access Policy*. Retrieved from http://www.cihr-irsc.gc.ca/e/46068.html

Digital Curation Centre (n.d.) *How to Cite Datasets and Link to Publications*. Retrieved from http://www.dcc.ac.uk/resources/how-guides/cite-datasets

Digital Curation Centre (n.d.) *How to License Research Data*. Retrieved from http://www.dcc.ac.uk/resources/how-guides/license-research-data

DMP Tool (2014) *Data Management General Guidance*. Retrieved from https://dmptool.org/dm_guidance

Krier, L., Strasser, C. A., (2014). Data management for libraries: A LITA guide. Chicago: ALA TechSource.

Natural Sciences and Engineering Research Council of Canada (2014) *Responsibilities and Accountability*. Retrieved from http://www.nserc-crsng.gc.ca/Professors-Professeurs/FinancialAdminGuide-GuideAdminFinancier/Responsibilities-Responsabilites_eng.asp

Social Sciences and Humanities Research Council (2014). *Research Data Archiving Policy*. Retrieved from http://www.sshrc-crsh.gc.ca/about-au_sujet/policies-politiques/statements-enonces/edata-donnees_electroniques-eng.aspx

UK Data Archive. (n.d.) *Create and Manage Data: Consent and Ethics*. Retrieved from http://www.data-archive.ac.uk/create-manage/consent-ethics

U.K. Data Archive (n.d.) *Create and Manage Data: Copyright.* Retrieved from http://www.data-archive.ac.uk/create-manage/copyright

UK Data Archive. (n.d.) *Create and Manage Data: Planning for Sharing.* Retrieved from http://www.data-archive.ac.uk/create-manage/planning-for-sharing

University of Minnesota Libraries (n.d.) *Copyright and Ethical Use of Data.* Retrieved from https://www.lib.umn.edu/datamanagement/copyright

University of Minnesota Libraries (n.d.) *Sharing Your Data.* Retrieved from https://www.lib.umn.edu/datamanagement/datacenters

University of Oregon Libraries (n.d.) *Sharing Data.* Retrieved from http://library.uoregon.edu/datamanagement/sharingdata.html