

Data Camp Lesson Plan Outline

These materials were used for the 3-week instruction camp on research data management held at UWM Libraries in 2016. The camp was run by the Data Services Librarian and intended for other librarians to become more comfortable with data management concepts. The topics covered include: data management best practices, data management plans, and data repositories. Each 1-hour session consisted of 2 hands-on exercises. The first 3 pages of this document provide the general outline for each session, with the exercise handouts making up the rest of the document.

Week 1 - Data Management 101

- This week's goal is to cover background on the basic practices of good data management.
- Exercise 1
 - Given the scenario, have participants brainstorm everything they can think of that might go wrong with the data/digital content.
 - Potential problems:
 - Data errors
 - Lost paper sheets
 - Lack of consistency in spreadsheet file formatting
 - Access by unauthorized individuals
 - No plan for old data
 - General timing:
 - 10 minutes (small groups): Break into small groups and identify as many potential risks to the data as they can.
 - 5 minutes (whole group): Write the master list of risks on the board.
 - 10 minutes (small groups): Have each group choose a particular risk and come up with a plan to reduce that risk.
 - 10 minutes (whole group): Share out and discuss plans. Offer feedback.
 - Acknowledgements:
 - Many thanks to Josh Bishoff (University of Minnesota) for inspiration for this exercise. A copy of the original exercise can be found with the Midwest Data Librarians Symposium materials here: http://dc.uwm.edu/mdls/2015/teaching_workshop/3/
- Exercise 2
 - [Exercise 2 builds on Exercise 1 by repeating the process for the participant's individual data. This does not have to be as thorough as part 1.]
 - Have participants think of a recent project they worked on that generated digital content. Brainstorm a handful of things that could have gone wrong (or did go wrong!) with the digital content for that project. Have the participants pick one problem and make a plan for how to avoid this problem in the future.
 - General timing:
 - 10 minutes (individual): Brainstorm and plan.
 - 5 minutes (whole group): Share out plans.

Week 2 - DMPs

- This week's goal is to be able to discuss the main points of a NSF DMP and to have the participants write a general DMP for their own content. These exercises build off of the good practices discussed in week 1.
- Exercise 1
 - Share outline of a generic NSF DMP and offer the scenario. Have participants break into groups (5 groups ideal, though less is okay), assign each group a section of the DMP template, and brainstorm what kind of information should go in the assigned portion.
 - General timing:
 - 5 minutes (whole group): Review DMP history and walk through main parts of the generic NSF DMP requirement. Assign each group a DMP section.
 - 10 minutes (small groups): Brainstorm ideas for the types of information to include in a particular DMP section.
 - 10 minutes (whole group): Have groups report out what they think should go in each DMP section. Offer feedback.
- Exercise 2
 - [Exercise 2 builds on Exercise 1 by repeating the process for the librarian's individual content. This does not have to be as thorough as part 1.]
 - Using a slightly looser framework, have the participants come up with a simple DMP for their own content. An example DMP/README documentation file can be provided.
 - Good options for personal content include:
 - Committee minutes and files
 - Project involving multiple people
 - A large group of the participant's own digital files
 - General timing:
 - 10 minutes (individual): Brainstorm and plan.
 - 5 minutes (whole group): Share out plans.

Week 3 - Data Repositories

- The goal of this week is to expose participants to new data repositories in their liaison fields and have them learn how to evaluate a research data repository.
- Exercise 1
 - Divide participants into two groups and have one group brainstorm what makes a good repository for finding research data and the other brainstorm what qualities to look for in a repository for depositing research data.
 - General timing:
 - 2 minutes (whole group): Briefly review repository landscape, break participants into 2 groups and assign each a topic.
 - 10 minutes (small groups): Brainstorm what qualities make a good data repository for the group's assigned purpose.
 - 10 minutes (whole group): Have groups report out repository attributes. Offer feedback.
- Exercise 2
 - [This exercise requires participants have computers]
 - This exercise will allow participants to actually review data repositories. Participants can choose a repository for evaluation from the master re3data.org list, from a journal's recommended repository list, or a general repository. Participants may wish to look at a repositories associated with their liaison department(s).
 - General timing:
 - 2 minutes (whole group): Hand out and briefly review evaluation rubric. Participants can either use this or the criteria from Exercise 1 to evaluate repositories.
 - 5 minutes (whole group): Review re3data.org as a comprehensive list and suggest other repository/repository lists for starting points in choosing what to evaluate. At least a couple participants may wish to look at some of the big general repositories.
 - 15 minutes (individual): Have participants choose a repository to review and let them how good it is using the brainstormed or provided rubric.
 - 5 minutes (whole group): Have people report back on what they liked and didn't like about the repositories they visited. Were there any repositories in particular that people really liked and would recommend to patrons?

Exercise

The Library gathers statistics (number/type of questions per hour, etc.) for the reference and welcome desks each day. Tallies are done by hand on a paper form, with one form being used for each desk each day across the multiple librarians and student workers working different shifts. This generates roughly 60 forms per month (1 per desk per day), which are compiled by one of the reference librarians at the end of each month into an Excel document. This document is then placed in a public folder on the Library's staff server and emailed to the relevant people across the library to determine staffing, etc.

What could go wrong?

Exercise

Think of a recent project you worked on that generated digital files. Think of one thing that could have gone wrong or did go wrong. How can you prevent this issue from arising on your next project?

A NSF DMP should include

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.

Scenario

An LIS researcher studies how patrons use the library website to search for content. The researcher both observes patrons in real time as well as conducts surveys. The researcher is open to sharing the final data from the project.

Exercise

Pick one DMP section and brainstorm the kind of information that belongs in that section. This can be either specific information from the scenario or what you think the researcher should do based on last week's session.

DMP Outline for Your Digital Files

What content do you have? Where is it stored and backed up?

What documentation does your content have or need?

How is your content organized? Do you use a naming convention?

Who has access? Do you need to restrict access for any reason? How will you do this?

What happens to the content once the project is finished?

Data Camp Week 2 – Exercise 2 Example

This document describes the general organization and naming conventions for the LSA committee files. Please follow stated conventions when adding content to this folder and its subfolders. This will help future LSA Committee members find files with minimal effort.

FOR NEW LSA COMMITTEE MEMBERS

Refer to content in the "NewCommitteeInfo" folder for introductory information on committee responsibilities.

FOLDER ORGANIZATION

Content is organized into the following folders:

- Bylaws
- Events
 - DiversityEvents
 - FundraisingEvents
 - HolidayParties
 - LengthOfServiceAwards
 - MembershipEvents
 - MoebiusAwards
 - Picnics
- Finances
- MeetingMinutes
- Membership
- Miscellaneous
 - Photos
 - Surveys
 - VolunteerRecruiting
- NewCommitteeInfo

FILE NAMING

Always start file names with a four-digit year. This should also be done for subfolders, unless listed above.

ex: "2006_DiversityCelebration", "2011_IncomeAndExpenses", "2002_Bylaws", etc.

Files requiring more date information should use the convention YYYY-MM-DD

ex: "2013-07-26_LSAPicnic", "2008-10-15_LSAMeetingMinutes", etc.

Follow naming patterns of existing and related files/folders whenever possible.

Avoid spaces in names, as well as the following characters: " / \ : * ? ' < > [] & \$

Label draft and final versions separately, preferably using "FINAL" at the end of the file name of the final version.

What Makes a Good Data Repository?

Brainstorm ideas on one of two topics:

Topic 1 – What should users look for in a data repository for finding data?

Topic 2 – What should users look for in a data repository for depositing their research data?

Is This a Good Data Repository?

1) Pick a repository

Find a subject-specific repository on one of these lists:

Re3data repository list

<http://www.re3data.org/browse/>

Scientific Data recommended repositories

<http://www.nature.com/sdata/data-policies/repositories>

Biosharing database list

<https://biosharing.org/databases/>

Or pick a general repository:

Figshare (general)

<https://figshare.com>

Zenodo (general)

<http://zenodo.org/>

Dryad (biology)

<http://www.datadryad.org/>

ICPSR (social science)

<http://www.icpsr.umich.edu/>

2) Evaluate repository using the provided rubric or the rubric from Exercise 1

Evaluating a data repository

Basics:

- URL for public access
- Information for depositors
- What organization is behind the repository?

Scope:

- Specialization in a particular area, subject, or format
- If specialized, any notable services for the data type of focus

Metadata and indexing:

- Metadata describing data in catalog record (e.g., any standard used?)
- Search features
- Metadata expected of the depositor

Deposit policies and workflows:

- Deposit size limits
- How are deposits initiated (i.e., point of contact, web form, etc.)
- Terms of deposit
- Are deposits open to researchers or only those with a particular affiliation?
- Are deposits automatically accepted or does the repository make a selection decision?

Fees for data deposit:

- One-time or ongoing? Do fees vary?
- What if any role does the depositor have in ongoing management of the data?

Review procedures:

- Review of data (incl. for confidential data)
- Review of documentation

Preservation features:

- Preservation/curation features
- File format conversion
- Time commitment for preservation and access

Access policies and procedures:

- Access restrictions
- Any fees for data use/download?
- Terms of use
- Ability to provide restricted access to confidential data
- Bulk access services (e.g., APIs)

Tracking and citation:

- Citation generated for end users
- Persistent identifier used
- Download statistics provided

Based on MIT Libraries' Repository Comparison Chart Template (CC BY-NC)

http://libraries.mit.edu/data-management/files/2014/05/RepositoryComparisonChartTemplate_Generalized_20161.rtf