Data Management Planning: Informing Research Data Management in Academic Libraries Today and the Skills Needed for Future Librarians

Tess Grynoch, Sarah Stevenson, Michael Smit

Dalhousie University
Halifax, Canada

Abstract
Academic libraries in Canada are taking the lead in research data management (RDM) as they strive to assist researchers in addressing Canada's move to join the global trend of adding data dissemination and preservation requirements to public research funding. The transition to include academic research data in collections of local scholarly communications requires a substantial expansion of traditional library services. Academic libraries are addressing the addition of RDM services in three primary ways: building on existing expertise within their systems, participating in national and international collaborations and training opportunities, and looking to library and information science (LIS) programs to supply skilled RDM graduates. While managing data is currently a component of LIS education, this skill has only recently become emphasized within the curriculum, and even recent graduates may not have put this skill to use in library settings where working with information and knowledge is the norm. This paper will detail the experiential learning opportunity of a Master of Library and Information Studies student who applied data management skills learned in the classroom to a real-life setting. The student created a data management plan for a university researcher in the sciences by completing a data inventory, collecting information in interviews, reviewing related literature, and applying class-based learning. The resulting understanding of best practices in RDM and the experience of creating this DMP were used to inform RDM best practices and procedures at the Dalhousie University Libraries and the development of future RDM courses and training opportunities at Dalhousie University's School of Information Management, and may similarly inform other academic libraries and LIS/information schools.

Introduction
Over the past decade, research data dissemination and preservation requirements and expectations (often from funding bodies) have pushed researchers to retain, document, preserve, and share their scholarly work, including research data[1]. Academic libraries are rising to the challenge of supporting their research community by providing research data management (RDM) services. Academic libraries are addressing the addition of RDM services in two ways, by applying existing expertise within the library and looking to library and information studies/science (LIS) programs to supply a skilled RDM workforce. Expertise from different libraries is also being harnessed to produce national services and resources and formal/informal communities of practice (e.g., Portage in Canada, Research Data Access & Preservation from ASIS&T). Data management, currently a component of the LIS curriculum, has recently gained stronger emphasis in reaction to RDM, big data, and other trends within the library field and society as a whole.

Data has been a component of ALA-accredited degree programs for decades, but the skills to effectively manage data require additional investigation and study beyond the core curriculum. This paper details the experience of a directed studies course (the authors are, respectively, the student and the two course supervisors) exploring research data management. The course was built around three concurrent projects addressing three different challenges of RDM and RDM training: the development of a RDM course at Dalhousie's School of Information Management, the ongoing

[1] While these pressures exist internationally, and in some countries are more stringent, this paper is focused on US/Canada, particularly Canada.

development of research data services at Dalhousie Libraries, and the creation of a data management plan for a science researcher at Dalhousie University. Each project was collaboratively defined at a high-level but led and executed by the student, and each project informed the other two projects.

Curriculum Project: RDM in Teaching

Dalhousie's School of Information Management currently requires students to complete a data management course that provides a basic overview of analyzing, manipulating, and managing data. RDM is introduced in one of the modules and one of the assignments during the course. To further develop the RDM skill set, the School is creating an elective course dedicated to the subject. For this project, two sets of information were reviewed. First, an examination of the state-of-the-art in academic and grey literature to create a list of required and recommended readings for this proposed course, annotated with summaries and the reason for the recommendation. Second, an environmental scan of RDM courses at ALA-accredited schools to identify current best practices in teaching RDM. These two large sets of information were synthesised into a set of proposed topics for the course (Table 1).

Although data is not explicitly mentioned in the curriculum requirements for information schools' American Library Association (2015) accreditation, many schools provide data courses and courses on research data management. An environmental scan of courses with research data management content was performed to determine possible topics for the course and syllabi were collected when available and applicable (Table 2). A list of assignments in use (with brief descriptions) was also generated, based on the available information. It was noted that many schools provide data mining courses, a topic of relevance to data management but not explicitly within scope of this scan (Dalhousie's SIM offers a related course, Business Analytics, which is cross-listed with the School of Business). Future scans should consider including other data-related courses in their scope, as a broad set of data skills is beneficial to RDM professionals.

The class topics found in these syllabi were synthesized with the working document on "Librarians' competencies profile for research data management" compiled by the Joint Task Force on Librarian's Competencies in Support of E-Research and Scholarly Communication's (2015). Since there are 12 weeks of classes, the topics and skills were synthesized into 12 modules that also took into consideration where topics could be removed or consolidated due to coverage in pre-requisite classes in the program. The 12 modules were then used to create an annotated reading list. Plans are currently in place to implement the course in fall 2017.

1. A short review (defining data, types of data, and data formats) and why data needs to be managed and shared
2. Data lifecycle and models, data sharing and discovery, assessing data repositories, and citing data you find
3. Data management plans: Their content, tools to help build them, and the data interview
4. The research practices and data management needs of science and health researchers
5. The research practices and data management needs of social science and humanities researchers
6. Data sharing and reuse: Practice vs. Policy
7. Metadata for data: Metadata crosswalks and how to explain metadata to non-librarians
8. Software and hardware for data management
9. Legal implications to data sharing: Addressing sensitive data and intellectual property concerns
10. Assessing data: Collection development and digital preservation
11. The price of repositories and data services: Sustainable and economic models
12. Future of data management: Semantic web, linked data, other emerging data trends, and job prospects in this field

Table 1. List of proposed topics for a research data management course synthesized from academic & gray literature and current best practices.

Institution	Course(s)	Syllabi Available
University of Arizona	Data Management in Healthcare Systems	No
University of British Columbia	Research Data Management for Information Professionals	Yes
University of California, Los Angeles	Data Management and Practice; Data Curation and Policy	Yes
University of Illinois	Foundations of Data Curation; Research Data Analysis and Management in LIS; Research Problems in Data Curation; Scientific Data Policy Seminar	Yes
University of North Carolina, Chapel Hill	Data Curation and Management	Yes
University of Pittsburgh	Research Data Management; Research Data Infrastructure	Yes
Simmons College	Scientific Research Data	No
University of Texas, Austin	Introduction to Scientific Data Informatics	Yes
University of Washington	Fundamentals of Data Curation	No
Western University	Special Topic: Research Data Management	No
University of Wisconsin, Madison	Digital Curation	Yes
University of Toronto	Data Librarianship	No

Table 2. Results of environmental scan for relevant research data management courses in ALA accredited institutions.

Librarian's Guide Project: RDM in Service

Dalhousie Libraries accelerated their development of data management services in 2014 and have since developed a data management services page on the library's website, a research data management LibGuide, and a deployment plan for a institutional data repository built on Dataverse. Dalhousie Libraries is also preparing for the expected data management plan (DMP) requirements from Canadian public research funders (Government of Canada, 2015). One area of particular importance is growing capacity within the library system to provide personal service and expertise to faculty members seeking to develop research data management plans. Centralized initiatives like websites, guides, and repositories do not require the same level of general expertise and training as distributed initiatives where subject liaison librarians are meeting individually with faculty members.

This project involved crafting materials to support the Dalhousie Libraries as they engaged in research data management conversations with university faculty. To establish requirements and expectations, librarians leading Dalhousie's RDM initiatives were consulted. They identified five core requirements:

- A checklist of steps and key pieces of information to collect, adapted for different stages of research (from research planning and design to post-data-collect). They were able to supply a set of questions about the data practices of researchers who wished to archive their data.
- Not text heavy: brief and to the point.
- Something that can be provided to faculty members (not targeted specifically at librarians)
- Specific questions and guidelines for how to conduct a data/reference interview in this area, including examples of different types and formats of data.
- Because not all librarians are comfortable with providing research data management support, the guidelines should be approachable for librarians with a wide range of experience and comfort.

Given the requirement to cover various stages of research, two separate scenarios were considered: one intended to guide a data interview with researchers wanting help managing project data who approached the library prior to beginning data collection, and a similar-but-different set of

questions for researchers wanting help preserving or sharing their data following their collection. The current organization the library's data services page aligns with these two scenarios. The request for specific questions guided the remainder of this project: questions in the form of an interview guide are the primary deliverable. As in the case of the curriculum project, the questions were developed based on a review of current best practices and of academic and grey literature on research data management.

Questions were built from the existing series of questions regarding data archiving in use, the Portage (2016) DMP (Data Management Plan) Assistant, and the Purdue University Libraries' (2011) Data Management Plan Self-Assessment Questionnaire and Data Curation Profiles Toolkit: Interview Worksheet (Carlson, 2010). One of the concerns about Purdue's Data Curation Profiles is the length of time needed for the interview; because researchers and librarians are both busy, it is important that the interview be as short as possible (while also being thorough) (Zhang, Zilinsky, Brandt, & Carlson, 2015).

The questions were divided into two sets of questions: one to be sent to researchers prior to an in-person meeting, and one to help guide the meeting itself. A pre-meeting set of questions can help researchers understand the kind of things they'll need to think about when research data management planning, and their responses can help librarians focus their preparation for the interview, including identifying appropriate resources and refining follow-up questions. It also reduces the length of the interview itself, off-loading content from a synchronous meeting to asynchronous communication. This format for the questions is modeled on Research by Appointment procedures and the theory of implementing current skills or practices present in the libraries, such as reference, towards data services (Cox & Corrall, 2013; Huwe, 2013; Mclure et al., 2014). When using the pre-meeting questions is not an option, or not preferred (or completed) by the researcher, the two set can be combined into a single interview.

Each of the synthesized questions is augmented with explanatory notes as required, explaining the purpose and goals of the question, and the type of information expected in response. The questions are presented in the form of a questionnaire for the pre-meeting questions, including a short generic message to the researcher, and in the form of an interview guide for the librarian conducting the in-person meeting. The resulting guides were iteratively shared and refined by the authors, librarians responsible for RDM at the Dalhousie Libraries, and the researcher in the following case study. The completed questionnaire / interview guide is provided in Appendix A.

Creating a Data Management Plan: RDM in Practice

This project involved a case study, completing a data inventory and conducting a research data management interview with a researcher interested in receiving data services from the Dalhousie Libraries. As mentioned, this project informed the previous two projects, but was also informed by the previous two projects.

Dr. Anna Metaxas is a marine larval and deep sea ecologist at Dalhousie University whose projects include field experiments, lab experiments, computer modeling, and capturing hours of video of marine ecosystems from her research cruises. Her main data management concern was the loss of videos that were collected at great time and expense, and have substantial potential value to future researchers seeking to study changes in marine environments, particularly deep sea ecosystems. These videos are stored on two different tape formats and various portable hard drives. This project was built around creating a data management plan, and guided by a video data inventory documenting the video currently held by Dr. Metaxas, as well as a data interview conducted using the interview guide created in the previously-described project. During continued use, we anticipate further refinements to the guidelines.

The data inventory counted the number of videos stored in tape formats, and recorded metadata from the tape cases and tapes themselves, in preparation for an eventual digitization process. It also collected the directories from the hard drives which held digital videos using command line tools, which were then cleaned to extract various pieces of metadata from the file names.

The data interview was conducted using a single set of questions merged from the pre-meeting and data interview question sets, due to familiarity with the data based on the data inventory and the preferences of Dr. Metaxas'. Questions were skipped if they could be answered based on the results of the data inventory. While a data inventory can be a time-consuming undertaking, it is a valuable way to collect accurate information about the data to be managed, and it reduces the length of the interview and the corresponding burden on the researcher. In cases where a complete data inventory is not feasible, reviewing the researcher's public profiles (websites, Google Scholar) and the policies of any funders identified (including agencies and networks) can also help answer questions and reduce the length of the interview. This initial interview with a thoughtful and helpful faculty ally helped identified questions that needed further clarification, and some changes to the order of the questions.

Using the results of the interview and the data inventory, augmented with additional examination of best practices and discipline-specific standards, a data management plan was created for the lab. One challenge identified was the variable data management procedures by students within her lab. She expressed concerns of data leaving with students, lack of clarity of data submitted to her by students before they leave, and data safety concerns such as file backup copies being kept on hard drives with files that were in active use. To address these concerns, a short document was created to translate the data management plan into practical, day-to-day best practices for students in the lab.

One important lesson from this case study is related to the first project, teaching RDM in LIS courses. The current draft syllabus includes a project to create a data management plan for interested researchers. The amount of work involved in this case study suggests this plan needs revision to define the scope more precisely. Rather than a DMP for a lab, students could craft a DMP for a single project; or, a group project approach could be used to make crafting a DMP for a lab more feasible while avoiding potentially differing and fragmented DMPs within a research group. We also suggest a scenario where information management students write data management plans for and with graduate research students. This cross-pollination approach could help graduate students on the front lines of research understand, internalize, and adopt data management practices. Graduate students may also be easier to recruit to this task than principle investigators.

Conclusion

These three projects are a reasonable proxy for the role of many academic librarians with RDM responsibilities, considering teaching, service, and practice. These three areas are complementary: the RDM principles gleaned from the potential course readings informed the guidelines created, the guidelines were used to compose the data management plan, conducting a data interview resulted in modifications to the guidelines, and the creation of the data management plan determined the feasibility of such a project within the proposed RDM course. This directed studies course provided an opportunity to expand the knowledge and experience of a student while contributing to the current state of research data management practice, at Dalhousie University in particular and the academic community in general. It also provided exposure to the true landscape of RDM: iterative processes, shifting requirements, evolving technologies, time-strapped stakeholders, and never enough resources. It helps students recognize that we cannot let the drive for perfection prevent us from producing and sharing something that is good. RDM provides a rich landscape for experiential learning designed to help students understand how to assess, understand, and meet the needs of a diverse group of stakeholders. RDM is one current challenge, and there will be others in the future as everyone adapts to the changes in academic endeavours that are driven by technology. As we strive to train students in RDM, we should ensure we continue to equip students with the critical-thinking and problem-solving skills to tackle not just RDM, but the various changes they will encounter in their careers.

References

American Library Association. (2015). Standards for accreditation of master's programs in library and information studies. Retrieved from http://www.ala.org/accreditedprograms/sites/ala.org.accreditedprograms/files/content/standards/Standards_2015_adopted_02-02-15.pdf

Carlson, J. (2010). The Data Curation Profiles Toolkit: Interview Worksheet. *Data Curation Profiles Toolkit.* Paper 3. doi: http://dx.doi.org/10.5703/1288284315652

Cox, A. M., & Corrall, S. (2013). Evolving academic library specialties. *Journal of the American Society for Information Science & Technology, 64*(8), 1526–1542. doi: 10.1002/asi.22847

Government of Canada. (2015). Draft Tri-Agency Statement of Principles on Digital Data Management - Science.gc.ca. Retrieved from http://www.science.gc.ca/default.asp?lang=En&n=83F7624E-1

Huwe, T. K. (2013). Data discovery and data curation going hand in hand. *Computers in Libraries, 33*(3), 17–19. Retrieved from http://www.infotoday.com/cilmag/

Joint Task Force on Librarian's Competencies in Support of E-Research and Scholarly Communication. (2015). Librarians' competencies profile for research data management. Retrieved from https://docs.google.com/document/d/15Kqg6w6AFsdbbaxcRCx7DZ3OERnQk6E0UhQ4UeRu0Cc/edit

Mclure, M., Level, A. V, Cranston, C. L., Oehlerts, B., & Culbertson, M. (2014). Data curation: A study of researcher practices and needs. *Portal: Libraries and the Academy, 14*(2), 139–164.

Portage. (2016). DMP Assistant. Retrieved from https://portagenetwork.ca/

Purdue University Libraries. (2016, February 4). Data management plan self-assessment questionnaire. Retrieved from https://purr.purdue.edu/app/site/media/docs/dmp_self_assessment.pdf

Zhang, Zilinsky, Brandt, & Carlson. (2015). Assessing perceived usability of the data curation profile toolkit using the technology acceptance model. *The International Journal of Digital Curation, 10*(1), 48-67. Retrieved from http://repository.cmu.edu/cgi/viewcontent.cgi?article=1161&context=lib_science

Purdue University Libraries. (2016, February 4). Data management plan self-assessment questionnaire. Retrieved from https://purr.purdue.edu/app/site/media/docs/dmp_self_assessment.pdf

Appendix A.
Research Data Services (Interview) Guidelines for Dalhousie University Libraries
Developed by Tess Grynoch

This series of questions is meant to guide discussions with researchers on the subject of how they would like to manage their data. Please make clear to researchers that they are allowed to say they do not know the answer to any of the questions.

For a researcher about to embark on data collection:

Preliminary Questions
(Ideally answered before a data reference interview is conducted)

Message to researcher
Dear _____,

In order to be best prepare for our discussion about your data management plan on __(date)__, please fill out this short form with ten questions about your research and the data you are planning to collect. If you do not know the answer to a specific question at this time, please feel free to respond as such to the applicable question(s).

1. What kind of research do you do? (General research questions, approaches, methodologies, etc.)
   Outcome: This question will help inform which librarians may need to be consulted to gain a background understanding of the subject and may even be beneficial to have the consulted librarian(s) present during the data interview.

2. Are you the principal investigator?
   Outcome: If they are not the PI, the PI may need to be involved in getting permission for sharing the data, and/or ensure they are aware of the plan.

3. What kind of data do you plan to collect/create? (e.g. qualitative, quantitative, models, code, etc.)

4. What formats are involved? (e.g. transcripts, numeric data, etc.)

5. How much data will you collect and how fast will it grow? (in MB/GB/TB)
   Outcome: Service fees may apply for storing data sets greater than 0.5TB.

6. Do you have any grants that require archiving or/and sharing your data?

7. If data archiving and/or sharing is grant-mandated, what type of grant do you have?
   -Are you part of a research network?
   -Are there existing constraints or requirements placed on your data archiving and/or sharing? (Grants, journal data policies, research networks, etc.)?

8. May the data be shared? If so, with whom (fellow researchers, policy makers, public, etc.)? Are there specific restrictions on the data (legal, ethical, intellectual property)?
   -May whom with the data be shared?

9. Will you collect the primary data yourself, or will it be obtained from another source? (e.g. a government department)
   Outcome: When there are multiple owners for a data set, the data is more difficult to share and another institution may already be handling the preservation and sharing of the data.

10. How have you organized, stored, and shared your data in the past?
    Outcome: Easier to align best practices to habits they already have. This question also provides insight into their current understanding of data management.

Interview Questions
Librarian has copy of the answered questions from the initial question answering. Ask clarification questions.

1. How will you organize your data? Are you partial to a particular folder layout, file-naming structure, or versioning control?
   - (This may have arisen in how they organized data in the past)

   Outcome: Helps to know the number of estimated data files when coming up with numbering conventions.
2. What documentation files will be created (readme.txt, lab notebooks) to provide a description of the data and methodology of how the data was collected (and analyzed, if not raw data)? If any data analysis software or tools will be used, which ones?

3. Will you be using a metadata standard to document and describe your data? How will your metadata be generated and captured?

Understanding the data life cycle

(May help to draw a flow chart similar to Barsky, E. (2015). *Research data discovery manual*. Retrieved from http://library-escience-collab.sites.olt.ubc.ca/files/2014/07/UBCLibrary_CHHM_RDMP_Manual_DataFlowAndNeedsDiscoverySessionManual_20140425_v11.pdf)

4. Will other members of the research team and other collaborators need to access, modify, and contribute data throughout the project? If so, how?
5. How and where will your data be stored and backed up during the various stages of your research project? If your research project includes sensitive data, how will you ensure that it is securely managed and accessible only to approved members of the project?

Sharing and preserving the data

6. Who do you imagine as the audience for your research data?
7. Where would you like to deposit your data for long-term preservation and/or access? Do you want or need to share the data from the same deposit location?
   - If submitting to a journal, does the journal then place limitations on that data?
   - Have you considered the Dalhousie repository or a domain-specific repository?
   - If depositing in a data repository, things to consider:
     - (Refer to Criteria for Evaluating Data Repositories)
     - Ability to access the data at a secondary (mirror) site if the repository is off-line
     - Being able to see data usage statistics

8. How will the data be shared, if not through the location where your data will be stored?
9. What data will you be sharing and in what form? (e.g. raw, processed, analyzed, final)
    - (Part of this discussion stems from considerations of the target audience and how they might use the data)
    - Which data sets have long-term value to others?
10. What needs to be done to the data to make it preservation ready and/or shareable?
    - Consider open (non-proprietary), machine-readable file formats, ensuring file integrity, anonymization and de-identification, inclusion of supporting documentation.
    - Examples of open (non-proprietary), machine-readable file formats include:
        - Comma separated value (.csv) files for tabular data
        - Shape files (.shp) or ASCII ArcGIS files for GIS data;
        - Tagged image file format (.tif or .tiff) for raster data and photographs
        - Etc. (Any file format where the code needed to read the file is freely available and the data can be read by a computer. These two requirements ensure that the files can be easily converted to other formats.)
11. Have you considered what type of end-user license to include with your data?
    - Any restrictions that need to be placed on the data (Refer to restrictions listed in question 8 of the preliminary questions)
    - Permitting re-use of the data, with or without conditions?
    - Permitting re-distribution of the data, with or without conditions?
    - Permitting the creation and publication of derivatives from the data, with or without conditions?
    - Permitting others to use the data to develop commercial products or in ways that produce a financial benefit for themselves, either with or without conditions?
    - How would you like to be attributed for your data sets?
    - (Creative commons policy builder)
12. Who will be responsible for managing this project's data during and after the project and what are the major data management tasks for which they will be responsible?
13. How will responsibilities for managing data activities be handled if substantive changes happen in the personnel overseeing the project's data, including a change of Principal Investigator?
14. What resources do you have to manage your data (including technology that is able to read the data)? What additional resources will you require to manage your data? What do you estimate the overall cost for data management to be?
15. What steps will be taken to help audience for your research data know that your data exists?
    - (Advocacy question. Sell it! Anything written down will be primarily for the researcher)
    - Even if data is stored, and archived it won't be used if people can't find it…
    - Methods to help the research community know that your data exists:
        - Creating a data citation
        - Adding a unique DOI or handle
        - Advertising on your website
        - Citing in your own work
    - Benefits of "marketing" your research data:
        - Data is a research output so you should get credit
        - Attention being paid in granting agencies
        - Increase standing in research community

For a researcher who is finished their data collection and is looking to archive:

Preliminary Questions
(Ideally answered before a data reference interview is conducted)

1. What kind of research do you do? (General research questions, approaches, methodologies, etc.)
   Outcome: This question will help inform which librarians may need to be consulted to gain a background understanding of the subject and may even be beneficial to have the consulted librarian(s) present during the data interview.

2. Are you the principal investigator?
   Outcome: If they are not the PI, the PI may need to be involved in getting permission for sharing the data, and/or ensure they are aware of the plan.

3. What kind of data do you collect/create? (e.g. qualitative, quantitative, models, code, etc.)

4. What formats are involved? (e.g. transcripts, numeric data)
   Outcome: If the formats are proprietary, need to ask if that format is necessary or if they can be turned into non-proprietary formats to ensure sharing and longevity.

5. How much data do you have? (in MB/GB/TB)
   Outcome: Service fees may apply for storing data sets greater than 0.5TB.

6. Do you have any existing documentation files (readme.txt, lab notebooks) that provide a description of the data, file naming conventions, and methodology of how the data was collected?

7. Do you have an existing metadata files with a key or reference to each data field?

8. Is the data archiving and/or sharing mandated by a grant or journal submission?

9. If data archiving and/or sharing is grant-mandated, what type of grant do you have?
   -Are you part of a research network?
   -Are there existing constraints or requirements placed on your data archiving and/or sharing? (Grants, journal data policies, research networks, etc.)?

10. May the data be shared? If so, with whom (fellow researchers, policy makers, public, etc.)? Are there specific restrictions on the data (legal, ethical, intellectual property)?
    -May whom with the data be shared?

11. Where is the data currently stored? (e.g. hard drive, departmental server)

12. Did you collect the primary data, or was the data obtained from another source? (e.g. a government department) Or are there ownership restrictions based on funding requirements or publication requirements?
    Outcome: When there are multiple owners for a data set it makes it more difficult to share and another institution may already be handling the preservation and sharing of the data.

Interview Questions

1. Who do you imagine as the audience for your research data?
2. Where would you like to deposit your data for long-term preservation and/or access? Do you want or need to share the data from the same deposit location?
    - Have you considered the Dalhousie repository or a domain-specific repository?
    - If depositing in a data repository, things to consider:
        - (Refer to Criteria for Evaluating Data Repositories)
        - Ability to access the data at a secondary (mirror) site if the repository is off-line
        - Being able to see data usage statistics
3. How will the data be shared, if not through the location where your data will be stored?
4. What data will you be sharing and in what form? (e.g. raw, processed, analyzed, final)
    - (Part of this discussion stems from considerations of the target audience and how they might use the data)
    - Which data sets have long-term value to others?
5. What needs to be done to the data to make it preservation ready and/or shareable?
    - Consider open (non-proprietary), machine-readable file formats, ensuring file integrity, anonymization and de-identification, inclusion of supporting documentation.
    - Examples of open (non-proprietary), machine-readable file formats include:
        - Comma separated value (.csv) files for tabular data
        - Shape files (.shp) or ASCII ArcGIS files for GIS data;
        - Tagged image file format (.tif or .tiff) for raster data and photographs
        - Etc. (Any file format where the code needed to read the file is freely available and the data can be read by a computer. These two requirements ensure that the files can be easily converted to other formats.)
6. Have you considered what type of end-user license to include with your data?
    - Any restrictions that need to be placed on the data? (Refer to restrictions listed in question 10 of the preliminary questions)
    - Permitting re-use of the data, with or without conditions?
    - Permitting re-distribution of the data, with or without conditions?
    - Permitting the creation and publication of derivatives from the data, with or without conditions?
    - Permitting others to use the data to develop commercial products or in ways that produce a financial benefit for themselves, either with or without conditions?
    - How would you like to be attributed for your data sets?
    - (Creative commons policy builder)
7. Who will be the main point of contact or responsible party for this data?
8. What resources do you have to manage your data (including technology that is able to read the data)? What additional resources will you require to manage your data? What do you estimate the overall cost for data management to be?
9. What steps will be taken to help the audience for your research data know that your data exists?
    - (Advocacy question. Sell it! Anything written down will be primarily for the researcher.)
    - Even if data is stored, and archived it won't be used if people can't find it…
    - Methods to help the research community know that your data exists:
        - Creating a data citation
        - Adding a unique DOI or handle

        - Advertising on your website
        - Citing in your own work
    - Benefits of “marketing” your research data:
        - Data is a research output so you should get credit
        - Attention being paid in granting agencies
        - Increase standing in research community