**Writing a Data Management Plan:**

**Guidance, language, and resources**

**April 01, 2019  
Thea Atwood, Data Services Librarian  
tpatwood@umass.edu**

****

**Writing a Data Management Plan**

Guidance on Data Management Plans (DMPs) is often sparse. To help you better understand the data management plan section of writing an NSF grant, this worksheet breaks each section into discrete elements, and from there, each section is broken down into four parts:

1. Guidance from the NSF on writing the data management plan.
2. Deeper questions help guide you through the process
3. A sample reply (note: this is sample language only, you must create a DMP that relates to your work -- boilerplate language will not yield a robust plan for your work). All sample language comes from the Public Plans in the DMPTool (see below). Some language has been modified to reflect UMass resources or best practices.
4. Space for your notes as we work through the DMP.

Note that the guidance provided in this workbook follows the generic guidelines from the NSF. Directorates, offices, divisions, programs, or other NSF units may have specific guidelines. Please be sure to check if the solicitation you are applying to has their own specialized DMP requirements here: <https://www.nsf.gov/bfa/dias/policy/dmp.jsp>

If you have any questions, please reach out to us at the Data Working Group! [dwg@library.umass.edu](mailto:dwg@library.umass.edu)

**Resources and links:**

Connecting with local experts:

Guidance from the UMass Amherst Data Working Group: <https://www.library.umass.edu/services/research/rds/>

Email the Data Working Group directly: [dwg@library.umass.edu](mailto:dwg@library.umass.edu)

Writing a DMP:

DMPTool: <https://dmptool.org/>

Getting your DMP reviewed:

Email the Data Working Group and Thea & her team will review!: [dwg@library.umass.edu](mailto:dwg@library.umass.edu)

Or reach out to Thea directly: [tpatwood@umass.edu](mailto:tpatwood@umass.edu)

FAQs:

ResearchDataQ: <http://researchdataq.org/index.php>

NSF’s Data Management & Sharing FAQ: <https://www.nsf.gov/bfa/dias/policy/dmpfaqs.jsp>

Granular concepts around data management & sharing:

Standard data formats

Licensing data

Timeline for sharing data

Where to share data

Where to store data

*Uniquely identify data*: Digital Object Identifiers; available with deposit of Scholarworks!

Readings around RDM:

Goodman A, Pepe A, Blocker AW, Borgman CL, Cranmer K, Crosas M, et al. (2014) Ten Simple Rules for the Care and Feeding of Scientific Data. PLoS Comput Biol 10(4): e1003542. <https://doi.org/10.1371/journal.pcbi.1003542>

Hart EM, Barmby P, LeBauer D, Michonneau F, Mount S, Mulrooney P, et al. (2016) Ten Simple Rules for Digital Data Storage. PLoS Comput Biol 12(10): e1005097. <https://doi.org/10.1371/journal.pcbi.1005097>

Michener WK (2015) Ten Simple Rules for Creating a Good Data Management Plan. PLoS Comput Biol 11(10): e1004525. <https://doi.org/10.1371/journal.pcbi.1004525>

Strasser, C., Cook, R., Michener, W., & Budden, A. (2012). Primer on data management: what you always wanted to know. <https://www.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf>

**Public Data Management Plans used for this packet:**

|  |  |
| --- | --- |
| Section 0: | DMP dmpcurator, DMPTool, (2014). “A unified approach to preserving cultural software objects and their development histories.” *DMPTool.org.* <https://dmptool.org/plan_export/8273.pdf>  Li, K., (2014). “Doctoral Dissertation Research: An agent-based model of population changes in a vulnerable coastal environment.” *DMPTool.org.* <https://dmptool.org/plan_export/11276.pdf> |
| Section 1: | DMP dmpcurator, DMPTool, (2014). “Atmospheric CO2 concentrations, Mauna Loa Observatory, Hawaii, 2011 - 2013.” *DMPTool.org.* <https://dmptool.org/plan_export/8276.pdf> |
| Section 2: | DMP dmpcurator, DMPTool, (2014). “Atmospheric CO2 concentrations, Mauna Loa Observatory, Hawaii, 2011 - 2013.” *DMPTool.org.* [https://dmptool.org/plan\_export/8276.pd](https://dmptool.org/plan_export/8276.pdf)  Aronow, C., (2015). “Language endangerment and revitalization at the dialectal level: A current analysis of Salasaca Kichwa.” *DMPTool.org.* <https://dmptool.org/plan_export/15448.pdf> |
| Section 3: | Zappia, C. (2015). “Data management plans for the United States Migraine Study Protocol.” *DMPTool.org.* <https://dmptool.org/plan_export/17259.pdf> |
| Section 4: | Lin, J., Cruse, P., Fenner, M., & Strasser, C. (2014). Making Data Count: A Data Metrics Pilot Project. *UC Office of the President: California Digital Library*. Retrieved from <https://escholarship.org/uc/item/9kf081vf> |
| Section 5: | Lin, J., Cruse, P., Fenner, M., & Strasser, C. (2014). Making Data Count: A Data Metrics Pilot Project. *UC Office of the President: California Digital Library*. Retrieved from <https://escholarship.org/uc/item/9kf081vf> |

|  |
| --- |
| **Section 0: Roles and Responsibilities** |
| **Guidance from the NSF\*:**  “[The DMP] should outline the rights and obligations of all parties as to their roles and responsibilities in the management and retention of research data. It should also consider changes to roles and responsibilities that will occur should a principal investigator or co-PI leave the institution or project.”  \*This section is often not listed explicitly in the body of the guidance itself, but is often listed as part of the directions in filling out your DMP (i.e., this section is often buried in the instructions). |
| **Guiding Questions and Comments:**   * Outline the staff and organizational roles and responsibilities for implementing this data management plan. * Who will be responsible for data management and for monitoring the data management plan? * How will adherence to this data management plan be checked or demonstrated? * What process is in place for transferring responsibility for the data? * Who will have responsibility over time for decisions about the data once the original personnel are no longer available? |
| **Sample Reply:**  This data management plan will be implemented and managed by Riley Smith, under the project supervision of Principal Investigator Jordan Jones. The Scholarly Communication Department at the UMass Amherst Libraries will assist in transferring data to ScholarWorks@UMass, the Institutional Repository of UMass Amherst. The UMass Amherst Libraries will have long-term responsibility for the long-term storage needs of the data. All transferred data will be made publically accessible.  Should the PI leave the University of Massachusetts Amherst, Riley Smith will take responsibility for the storage and access of data, and transferring responsibility for the data to other lab personnel. |
|  |

|  |
| --- |
| **Section 1: Data produced** |
| **Guidance from the NSF:**  “The types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project.” |
| **Guiding Questions:**   * What data will be generated in the research? * What data types will you be creating or capturing? * How will you capture or create the data? * If you are using existing data, how will you obtain it? Do you have the appropriate agreements in place? * What is the relationship between the data you are collecting and any existing data? * How will the data be processed? * What quality assurance and quality control measures will you employ? |
| **Sample Reply:**  Air samples at Mauna Loa Observatory will be collected continuously from air intakes located at five towers – a central tower and four towers located at compass quadrants. Raw data files will contain continuously measured CO2 concentrations, calibration standards, references standards, daily check standards, and blanks. The sample lines located at compass quadrants were used to examine the influence of source effects associated with wind directions. In addition to the CO2 data, we will record weather data (wind speed and direction, temperature, humidity, precipitation, and cloud cover). Site conditions at Mauna Loa Observatory will also be noted and retained.  The final data product will consist of 5-minute, 15-minute, hourly, daily, and monthly average atmospheric concentration of CO2, in mole fraction in water-vapor-free air measured at the Mauna Loa Observatory, Hawaii. Data are reported as a dry mole fraction defined as the number of molecules of CO2 divided by the number of molecules of dry air multiplied by one million (ppm).  The final data product will be thoroughly documented in the open literature and in Scripps Institution of Oceanography Internal Reports. |
|  |

|  |
| --- |
| **Section 2: Data standards** |
| **Guidance from the NSF:**  “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).” |
| **Guiding Questions:**   * What file formats will you use for your data, and why? * What form with the metadata/documentation describing your data take? * How will you create or capture these details? * Which metadata standards will you use and why have you chosen them? (e.g., accepted domain-based standards, widespread usage) * What contextual details or documentation are needed to make the ata you capture or collect meaningful? |
| **Sample Reply:**  Surveys and field notes will be generated to .docx files and will be stored on an external hard drive. They will then be converted to a single PDF file each. PDF files are easily readable, are not easily retroactively edited, and consolidating all files will streamline the archiving process. TASCAM audio is recorded in uncompressed b-WAV format. WAV is the highest quality, least accessible and largest format, thus the audio will be analyzed in WAV, and all publicly accessible recordings will be archived in 356kbs MP3 format. The video/audio camcorder records in AVI, which is compatible FinalCut-Pro.  \*\*\*\*   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **Data type** | **Data format** | **Data stored** | **Shared format** | **Data stored** | | Survey and field notes | .docx | External hard drive in lab; UMass Box | .pdf | ScholarWorks@UMass | | TASCAM Audio | Uncompressed b-WAV | External hard drive in lab; UMass Box | 356kbs .mp3 | ScholarWorks@UMass | | Audio/video | AVI | External hard drive in lab; UMass Box | AVI | ScholarWorks@UMass | |
|  |

|  |
| --- |
| **Section 3: Policies for access and sharing** |
| **Guidance from the NSF:**  “Policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements.” |
| **Guiding Questions:**   * How will you make the data available? Include resources needed to make the data available: equipment, systems, expertise, etc. * When will you make the data available? * What is the process for gaining access to the data? * Will access be chargeable? * How long will the original data collector/creator/principal investigator retain the right to use the data before making them available for wider distribution? * Are there any embargo periods for political, commercial, or patent reasons? If so, give details. * Are there ethical and privacy issues? If so, how will these be resolved? * What have you done to comply with your obligations in your IRB protocol? * Who will hold the intellectual property rights to the data and how might this affect data access? |
| **Sample Reply:**  There are no requirements for making the results of the study private. All participants have signed waivers to be able to participate in the study, and to have their results shared, with the exception of any personal information. The USMAP group wishes to make their results available to the public, for anyone who is interested in taking rizatriptan or has questions about the drug, and for anyone within the USMAP group who was not directly part of the study.  The results of this study will not contain any personal information of the participants involved, nor will it contain any unique personal identifiers that could cause assumptions to be made about those involved. In order to assure that the results can be shared publicly, precautions have been taken to ensure that the participant's identities are kept anonymous.  The results of the study will be made public in the institutional repository at UMass Amherst upon completion of the project or at time of first paper publication, whichever comes first. Prior to sharing results publicly, data will be stored on a secure, encrypted, password protected laptop, which backs up to a department server and the UMass Amherst instance of Box.com on a weekly basis. The PI will also have access to data stored on the departmental server and Box.com, both of which are password protected. |
|  |

|  |
| --- |
| **Section 4: Reuse and redistribution of data** |
| **Guidance from the NSF:**  “Policies and provisions for re-use, re-distribution, and the production of derivatives.” |
| **Guiding Questions:**   * How do the policies from the previous section apply to reuse and redistribution of your data? * Will any permission restrictions need to be placed on the data? * Who is likely to be interested in the data? * What and who are the intended or foreseeable uses of the data? * What licenses will you apply to the data, if any? |
| **Sample Reply:**  All major data categories, as enumerated in Section 1, will be publicly available for review, evaluation, and  use as they are generated during the project and after its completion. Announcements about software and data availability will be made using a variety of channels (e.g., blogs, Twitter, email lists) targeting all interested stakeholder communities. All data will be licensed CC-BY to facilitate understanding of re-use. As data will be deposited in the institutional repository for the University of Massachusetts Amherst, data will be assigned a digital object identifier, and a recommended citation will be included as part of the descriptor of the data.  Similarly, all software products resulting from this project will be re-usable and redistributable both during the project and after its completion. The only restriction placed on redistribution of the software is that the copyright and license statement be kept intact as required by the Apache open source license. This software is expected to be of interest to publishers, data centers and repositories, individual researchers, and institutional administrators. |
|  |

|  |
| --- |
| **Section 5: Long-term access to data** |
| **Guidance from the NSF:**  “Plans for archiving data, samples, and other research products, and for preservation of access to them.” |
| **Guiding Questions:**   * What is the long-term strategy for maintaining, curating, and archiving the data? * Which archive/repository/database have you identified as a place to deposit data? * What procedures does your intended long-term data storage facility have in place for preservation and backup? * How long will or should data be kept beyond the life of the project? * What data will be preserved for the long-term? * What transformations will be necessary to prepare data for preservation or data sharing? * What metadata or documentation will be submitted alongside the data in order to make the data reusable? What other actions are necessary to make data reusable? (e.g., transformations) * What related information will be deposited? |
| **Sample Reply:**  In addition to managed in the community based GitHub repository, all major versions of the software  will be archived in UMass Amherst’s institutional repository (“ScholarWorks@UMass”) with persistent identifiers. Research data and records will be maintained for as long as they are of continuing value to the researchers and project collaborators. ScholarWorks@UMass has capabilities to manage, store, and share digital content, and provides persistent digital object identifiers, search interfaces, and tools for long-term data management. ScholarWorks is built off of the bePress architecture, which provides a secure, shared storage array with redundant backups.  Appropriate documentation will also be deposited with data. This contextual and project-level detail will help future researchers understand the scope and intent of the data, software used in data generation (including versions), persons involved in data generation, and will explain any shorthand (e.g., abbreviations, initials) or other elements that may be confusing to future users. If necessary, it will also describe organizational features of the dataset, such as file hierarchy.  As noted previously, data will be shared as comma separated values (CSV) and will not need to undergo any transformations before placing in ScholarWorks@UMass, as CSV is a highly interoperable data format. |
|  |