Data Management Plan
Self-Assessment Questionnaire

In general, your data management plan should address the following[1]:
- the types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;
- 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);
- policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
- policies and provisions for re-use, re-distribution, and the production of derivatives; and
- plans for archiving data, samples, and other research products, and for preservation of access to them.

Individual agencies or directorates may have additional items or more specific requirements for data management plans. More information about these requirements can be found at their respective websites (e.g. NSF - http://www.nsf.gov/bfa/dias/policy/dmp.jsp).

Describing the Research Data
For each data set, consider:
1. How will this data be generated and used in this project?
2. Describe the data set as completely as you can. Include information about the format, average size, volume, and/or estimated number of data files produced if possible.
3. Consider the lifecycles of this data set:
   a. What stages will the data pass through (e.g., raw, processed, analyzed, published)?
   b. What are your methodologies in each stage?
   c. What tools and instruments do you use?
   d. How much data will you generate and how fast will it grow?
   e. Who is involved (e.g., professors, lab techs, grad students)?
   f. How will this data set be managed (e.g., how and where will the data be stored and on what media)? How regularly, by whom, and how will data be backed up?
4. How will you identify and cover the costs of managing the data sets?

Data Standards
5. Are there any standard formats in your field for managing or disseminating the data sets you have identified (e.g., XML, ASCII, CSV, MySQL, netCDF)? If your format is proprietary rather than open, is this essential?
6. If there is not a standard format, how will you format your data so that others in your field will be able to make use of it?
7. Who on your team will have the responsibility for ensuring that data standards are properly applied and data are properly formatted?

Metadata Standards
Metadata is "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource"[2] (such as a data set). "A metadata record is a file of information which captures the basic characteristics of a data or information resource. It represents the *who, what, when, where, why* and *how* of the resource."[3]

8. How will metadata be generated and captured for each of your data sets?
9. Are you aware of any metadata standards specific to your field that could be used for your data sets (e.g., Dublin Core [DC], Resource Description Format [RDF], Federal Geographic Data Committee [FGDC], Directory Interchange Format [DIF], Ecological Metadata Language [EML], Minimum Information About a Proteomics Experiment [MIAPE], and the Data Documentation Initiative [DDI])?
10. If there is not a metadata standard, what metadata will you need to generate so that others in your field will be able to find, understand, and make use of your data?
11. Who are you research team will be responsible for ensuring metadata standards are followed?

Data Sharing
Funding agencies may require the sharing of data generated during the course of research conducted under a grant. Refer to the guidelines for your grant agreement for specific details.

12. Who would be the target audiences for your data sets (e.g., other researchers in field, researchers outside of field, policy makers, practitioners), and how would they use your data?
13. When will you share each of your data sets? You can use the table below if you have multiple data sets.

What Data: When:	Data Set #1	Data Set #2	Data Set #3	Other Data (if applicable)
Immediately after the data has been generated.				
After the data has been processed, normalized and/or corrected for errors.				
After the data has been analyzed.				
Immediately before publication.				
Immediately after the findings derived from this data have been published.				
Immediately after the funding for this project has expired.				
Within 6 months after the funding for this project has expired.				
Within 1 year after the funding for this project has expired.				
Other:				

14. Would you place any conditions on sharing your data with others (e.g., requiring some form of acknowledgement or attribution, forbidding for-profit use)?
15. If these data sets contain information gathered from or about human subjects or any other sensitive information, what steps will you take to ensure protection?
16. Describe any other ethical considerations to managing this data publicly.

Data Access

17. Are existing data repositories suitable for hosting your data sets and making them publicly available? What preparations would need to take place before you could transfer your data to the repository (e.g., reviewing the data to check for errors or ensure quality, ensuring compliance with grant funders' protocols or IRB requirements, obtaining sign off from stakeholders, gathering and/or reviewing relevant documentation to support the use, curation, or preservation of the data)?
18. If you don't have an existing repository, what methods, infrastructure, systems, mechanisms, or tools will you use to share your data?
19. What security measure will need to be provided in making these data sets available (e.g., permissions, restrictions, embargoes)?
20. Who will manage the security for these data sets and how?
21. Do you know how much it will cost to makes these sets available? How will you cover these costs?

Intellectual Property and Re-Use

Funding agencies may have varying approaches towards intellectual property, copyright and related issues. Please check with your funding agency and program officer(s) with any questions about specific requirements or questions about your data.

Intellectual property rights for data sets are subject to University policies. See http://www.lib.purdue.edu/uco/Resources/campus.html for a list of campus resources on copyright and intellectual property.

22. Who will own these data sets? Any other stakeholders need to be consulted before data sets are made available?
23. Will you permit the re-use of the data, either with or without conditions?
24. Will you permit the re-distribution of the data, either with or without conditions?
25. Will you permit the creation and publication of derivatives from the data, either with or without conditions?
26. Will you permit others to use the data to develop commercial products or in ways that produce a financial benefit for themselves, either with or without conditions?
27. How will the people who generated the data sets receive attribution for their work?

Data Archiving and Preservation

Digital Preservation can be defined as all of the "activities policies, strategies and actions to ensure access to reformatted and born digital content regardless of the challenges of media failure and technological change..."[4]

28. Which of your data sets have long-term value to others?
29. How will you ensure ongoing access beyond the life of the project?
30. What related information needs to be preserved with the data (e.g., software, reports, research papers, fonts, original bid proposal)?
31. How will you or the repository you are working with ensure that these data sets are able to withstand changes in or the obsolescence of the storage technologies?

Additional Resources

For additional information and resources on developing your data management plan please see the Purdue University Research Repository website at: http://research.hub.purdue.org/.

Purdue Librarians are available to assist you in identifying appropriate data and metadata standards, locating resources on developing policies for the sharing and re-use of data, find community-based data repositories, or finding additional resources on data management and curation issues. To arrange a consultation with a Purdue Librarian please contact your department's library liaison at:
http://www.lib.purdue.edu/rguides/instructionalservices/librarians.html.

This Self-Assessment Questionnaire was derived in part from the Data Curation Profiles (DCP) Toolkit, developed by the Purdue University Libraries in collaboration with the Graduate School of Library and Information Science at the University of Illinois Urbana-Champaign. The DCP Toolkit is designed to identify data management and curation needs associated with a specific data set. For more information about the DCP Toolkit please visit:
http://www.datacurationprofiles.org.

Notes & References

1. These five general areas to address in a data management plan originated from the National Science Foundation's Grant Proposal Guide: Chapter II - Proposal Preparation Instructions, NSF 11-1 January 2011.
http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp.

Please check with your funding agency and program officer(s) for more information or with any questions about specific requirements in managing, sharing or preserving data.

2. Zeng, Marcia and Qin, Jian *Metadata*. Version 1.3
 www.metadataetc.org/metadatabasics/overview.htm

3. Semerjian, Christopher J. "Metadata." In *Encyclopedia of GIS*. New York: Springer Science and Business Media LLC. 2008

4. American Library Association. "Definitions of Digital Preservation." ALA Annual Conference, Washington, D.C., June 24, 2007.
http://www.ala.org/ala/mgrps/divs/alcts/resources/preserv/defdigpres0408.pdf