Framework for Creating a Data Management Plan

This framework can be used as an outline in assembling data management plans to accompany grant applications. Note that some funders have page limits for data management plans—NSF limits plans to two pages.

See the University of Michigan ICPSR bibliography for additional readings germane to the elements of a data management plan.

Data Description (Highly recommended)

Provide a brief description of the information to be gathered -- the nature, scope, and scale of the data that will be generated or collected.

Why this is important

A good description of the data to be collected will help reviewers understand the characteristics of the data, their relationship to existing data, and any disclosure risks that may apply.

Example(s)

Access and Sharing (Highly recommended)

Indicate how you intend to archive and share your data and why you have chosen that particular option. Possible mechanisms for archiving and sharing include:

- Domain repository like ISSR (social science)
- Self-dissemination through a dedicated Web site that the research team will create and maintain. *If this option is chosen, it is recommended that the data producer arrange for eventual archiving of the data after the self-dissemination period terminates and specify the schedule for data sharing in the grant application.*
- Preservation with delayed dissemination. Under such an agreement the data producer makes an arrangement with a public data repository for archival preservation of the data with dissemination to occur at a later date, usually within a year.
- Institutional repositories. Institutional repositories at academic institutions have the goal of preserving and making available some portion of the academic work of their students, faculty, and staff. Note that not all IRs have the capacity to accept and curate data.

Why this is important

Sharing data helps to advance science and to maximize the research investment. A recent paper [link: http://deepblue.lib.umich.edu/handle/2027.42/78307] reported that when data are shared through an archive, research productivity increases and many times the number of publications result as opposed to when data are not shared.

Protecting research participants and guarding against disclosure of identities are essential norms in scientific research. Data producers should take efforts to provide effective informed consent statements to respondents, to deidentify data before deposit when necessary, and to communicate to the archive any additional concerns about confidentiality. (See Ethics and Privacy below.)
With respect to timeliness of data deposit, archival experience has demonstrated that the durability of the data increases and the cost of processing and preservation decreases when data deposits are timely. It is important that data be deposited while the producers are still familiar with the dataset and able to transfer their knowledge fully to the archive.
Example(s)

Will your data be free of direct and indirect identifiers? If not, how will you share your restricted data? Will special terms of use be required?
Example(s)

Indicate when the data will be made available to others.
Example(s)

**Metadata (Highly recommended)**

What types of metadata will you produce to support the data? Will a metadata standard be used?

**Why this is important**

Good descriptive metadata are essential to effective data use. Metadata are often the only form of communication between the secondary analyst and the data producer, so they must be comprehensive and provide all of the needed information for accurate analysis.

Structured or tagged metadata, like the XML format of the Data Documentation Initiative (DDI) standard, are optimal because the XML offers flexibility in display and is also preservation-ready and machine-actionable.
Example(s)

**Intellectual Property Rights (Highly recommended)**

Who will hold intellectual property rights for the data and other information created by the project?

Will these rights be transferred to another organization for data distribution and archiving? Will any copyrighted material (e.g., instruments or scales) be used? If so, how will the project obtain permission to use the materials and disseminate them?

**Why this is important**

In order to disseminate data, archives need a clear statement from the data producer of who owns the data. The principal investigator's university is usually considered to be the holder of the intellectual property rights for data the PI generates. Many archives do not ask for a transfer of rights but instead just request permission to preserve and distribute the data. Copyright may also come into play if copyrighted instruments are used to collect data. In these cases, data producers should initiate discussions with archives in advance of data deposit.
Example(s)
Ethics and Privacy (Highly recommended)
If applicable, how will you handle informed consent with respect to communicating to respondents that the information they provide will remain confidential when data are shared or made available for secondary analysis?

Why this is important
Protection of human subjects is a fundamental tenet of research and an important ethical obligation for everyone involved in research projects. Disclosure of identities when privacy has been promised could result in lower participation rates and a negative impact on science.
Example(s)

If applicable, what are your plans to obtain IRB approval?
Example(s)

Are there legal constraints (e.g., HIPAA) on sharing data?
Example(s)

If applicable, how will you manage disclosure risk in the data to be shared and archived?
Example(s)

Format (Highly recommended)
Specify the anticipated submission, distribution, and preservation formats for the data and related files (note that these formats may be the same).

Why this is important
 Depositing data and documentation in formats preferred for archiving can make the processing and release of data faster and more efficient.

Extensive community-based preservation practice has demonstrated that relying upon preservation formats that are platform-independent and non-proprietary is a sustainable approach to ensuring long-term access.
Example(s)

Archiving and Preservation (Highly recommended)
How will you ensure that data are preserved for the long term?

Why this is important
Digital data need to be actively managed over time to ensure that they will always be available
and usable. This is important in order to preserve and protect our shared scientific heritage as technologies change. Depositing data resources with a trusted digital archive can ensure that they are curated and handled according to good practices in digital preservation.

Example(s)

**Storage and Backup (Highly recommended)**

How and where will you store copies of your research files to ensure their safety? How many copies will you maintain and how will you keep them synchronized?

**Why this is important**

Digital archives have used archival storage strategies that rely upon geographically-distributed replication of preservation copies successfully for decades. A key to success is to continually incorporate the capabilities of current storage management and related technologies to ensure the longevity of preserved digital content.

Example(s)

**Security (Optional)**

How will you ensure that the data are secure?

**Why this is important**

Security for digital information is important over the data life cycle. Raw research data may include direct identifiers or links to direct identifiers and should be well-protected during collection, cleaning, and editing. Processed data may or may not contain disclosure risk and should be secured in keeping with the level of disclosure risk inherent in the data. Secure work and storage environments may include access restrictions (e.g., passwords), encryption, power supply backup, and virus and intruder protection.

Example(s)

**Responsibility (Optional)**

Who will act as the responsible steward for the data throughout the data life cycle?

**Why this is important**

Typically data are owned by the institution awarded a Federal grant and the principal investigator oversees the research data (collection and management of data) throughout the project period. It is important to describe any atypical circumstances. For example, if there is more than one principal investigator the division of responsibilities for the data should be described.

Example(s)

**Existing Data (Optional)**

Are there existing data with a focus similar to the data that will be produced? If so, list what they are and explain why it is important to collect new data.
**Why this is important**
This is important to include in a data management plan when the value of a new data collection comes from its relationship to existing data sources.
Example(s)

---

**Data Organization (Optional)**
Indicate how the data will be managed during the project, with information about version control, naming conventions, etc.

**Why this is important**
It is important to describe situations in which research data are in some way atypical with respect to how they will be organized. For example, some data collections are dynamically changing and version control is central to how the data will be used and understood by the scientific community.
Example(s)

---

**Quality Assurance (Optional)**
Specify how you will ensure that the data meet quality assurance standards.

**Why this is important**
Producing data of high quality is essential to the advancement of science, and every effort should be taken to be transparent with respect to data quality measures undertaken across the data life cycle.
Example(s)

---

**Budget (Optional)**
How will the costs for creating data and documentation suitable for archiving be paid?

**Why this is important**
How will the costs for creating data and documentation suitable for archiving be paid?
Example(s)

---

**Legal Requirements (Optional)**
Indicate whether any legal requirements apply to archiving and sharing your data.

**Why this is important**
Some data have legal restrictions that impact data sharing—for example, data covered by HIPAA, proprietary data, and data collected through the use of copyrighted data collection instruments. How these issues might impact data sharing should be described fully in the data management plan.
Example(s)
Selection and Retention Periods (Optional)
Indicate how data will be selected for archiving, how long the data will be held, and what your plans are for eventual transition or termination of the data collection in the future.

Why this is important
Data archives establish criteria for selecting data they commit to preserving. Producers of data should establish a clear case for the long-term value of each set of data produced.

Example(s)

Audience (Optional)
Describe the audience for the data you will produce.

Why this is important
The audience for the data may influence how the data are managed and shared—for example, when audiences beyond the academic community may use the research data.

Example(s)