Elements of a Data Management Plan

Element	Description
Data description	A description of the information to be gathered; the nature and scale of the data that will be generated or collected.
Existing data	A survey of existing data relevant to the project and a discussion of whether and how these data will be integrated.
Format	Formats in which the data will be generated, maintained, and made available, including a justification for the procedural and archival appropriateness of those formats.
Metadata (or data about data)	A description of the metadata to be provided along with the generated data, and a discussion of the metadata standards used. A metadata record is a file of information, usually presented as an XML document, which captures the basic characteristics of a data or information resource. It represents the ***who, what, when, where, why*** and ***how*** of the collected data.
Storage and backup	Storage methods and backup procedures for the data, including the physical and cyber resources and facilities that will be used for the effective preservation and storage of the research data.
Security	A description of technical and procedural protections for information, including confidential information, and how permissions, restrictions, and embargoes will be enforced.
Responsibility	Names of the individuals responsible for data management in the research project.
Intellectual property rights	Entities or persons who will hold the intellectual property rights to the data, and how IP will be protected if necessary. Any copyright constraints (e.g., copyrighted data collection instruments) should be noted.
Access and sharing	A description of how data will be shared, including access procedures, embargo periods, technical mechanisms for dissemination and whether access will be open or granted only to specific user groups. A timeframe for data sharing and publishing should also be provided.
Audience	The potential secondary users of the data.
Selection and retention periods:	A description of how data will be selected for archiving, how long the data will be held, and plans for eventual transition or termination of the data collection in the future.
Archiving and preservation	The procedures in place or envisioned for long-term archiving and preservation of the data, including succession plans for the data should the expected archiving entity go out of existence.
Ethics and privacy	A discussion of how informed consent will be handled and how privacy will be protected, including any exceptional arrangements that might be needed to protect participant confidentiality, and other ethical issues that may arise.
Budget	The costs of preparing data and documentation for archiving and how these costs will be paid. Requests for funding may be included.
Data organization	How the data will be managed during the project, with information about version control, naming conventions, etc.
Quality Assurance	Procedures for ensuring data quality during the project.
Legal requirements	A listing of all relevant federal or funder requirements for data management and data sharing.

Software Guide

Content Management System (CMS) is software that keeps track of every piece of content on your Web site, much like your local public library keeps track of books and stores them. Content can be simple text, photos, music, video, documents, or just about anything you can think of. A major advantage of using a CMS is that it requires almost no technical skill or knowledge to manage.

Data Documentation Initiative (DDI) is an effort to create an international standard for describing data (**metadata**). Expressed in XML, the DDI metadata specification now supports the entire research data life cycle. DDI metadata accompanies and enables data conceptualization, collection, processing, distribution, discovery, analysis, repurposing, and archiving. **http://www.ddialliance.org**

Drupal is an open source content management platform powering millions of websites and applications. It's built, used, and supported by an active and diverse community of people around the world. **http://drupal.org/**

Joomla is a content management system (CMS), which enables you to build Web sites and powerful online applications. Many aspects, including its ease-of-use and extensibility, have made Joomla the most popular Web site software available. Best of all, Joomla is an open source solution that is freely available to everyone. **http://www.joomla.org/**

Examples

Element	Description
Data description	A description of the information to be gathered; the nature and scale of the data that will be generated or collected. **Generic Example 1:** This project will produce public-use nationally representative survey data for the United States covering Americans' social backgrounds, enduring political predispositions, social and political values, perceptions and evaluations of groups and candidates, opinions on questions of public policy, and participation in political life. **Generic Example 2:** This project will generate data designed to study the prevalence and correlates of DSM III-R psychiatric disorders and patterns and correlates of service utilization for these disorders in a nationally representative sample of over 8000 respondents. The sensitive nature of these data will require that the data be released through a restricted use contract.
Existing data	A survey of existing data relevant to the project and a discussion of whether and how these data will be integrated. **Generic Example 1:** Few datasets exist that focus on this population in the United States and how their attitudes toward assimilation differ from those of others. The primary resource on this population, [give dataset title here], is inadequate because... **Generic Example 2:** Data have been collected on this topic previously (for example: [add example(s)]). The data collected as part of this project reflect the current time period and historical context. It is possible that several of these datasets, including the data collected here, could be combined to better understand how social processes have unfolded over time.
Format	Formats in which the data will be generated, maintained, and made available, including a justification for the procedural and archival appropriateness of those formats. **Generic Example 1:** Quantitative survey data files generated will be processed and submitted to the [repository] as SPSS system files with Data Documentation Initiative (DDI) XML documentation. The data will be distributed in several widely used formats, including ASCII, tab-delimited (for use with Excel), SAS, SPSS, and Stata. Documentation will be provided as PDF. Data will be stored as ASCII along with setup files for the statistical software packages. Documentation will be preserved using XML and PDF/A. **Generic Example 2:** Digital video data files generated will be processed and submitted to the [repository] in MPEG-4 (.mp4) format
Metadata	A description of the metadata to be provided along with the generated data, and a discussion of the metadata standards used. **Generic Example 1:** Metadata will be tagged in XML using the Data Documentation Initiative (DDI) format. The codebook will contain information on study design, sampling methodology, fieldwork, variable-level detail, and all information necessary for a secondary analyst to use the data accurately and effectively. **Generic Example 2**: The clinical data collected from this project will be documented using Clinical Data Interchange Standards Consortium (CDISC) metadata standards.
Storage and backup	Storage methods and backup procedures for the data, including the physical and cyber resources and facilities that will be used for the effective preservation and storage of the research data. **Generic Example 1:** [Repository] will place a master copy of each digital file (i.e., research data files, documentation, and other related files) in Archival Storage, with several copies stored at designated locations and synchronized with the master through the Storage Resource Broker.
Security	A description of technical and procedural protections for information, including confidential information, and how permissions, restrictions, and embargoes will be enforced. **Generic Example 1:** The data will be processed and managed in a secure non-networked environment using virtual desktop technology.

Generic Example 2: The main computer room that stores the program computer servers and network security equipment is protected by additional security features that require a separate authorization. Access to the data center is controlled by multiple biometric fingerprint readers. Any access to the computer by non-center persons is logged and requires a continuous escort by one of the center staff.

Generic Example 3: The secure server environment that hosts the database is located within a hardened data center at the [repository], and is governed by standard Medical information security guidelines including the Health Insurance Portability and Accountability Act of 1996 (Public Law 104-191), also known as HIPAA.

Responsibility

Names of the individuals responsible for data management in the research project.

Generic Example 1: The project will assign a qualified data manager certified in disclosure risk management to act as steward for the data while they are being collected, processed, and analyzed.

Generic Example 2: [repository] has two staff members whose responsibilities include the protection of HIPAA controlled information. The areas of responsibilities have been separated between the HIPAA Compliance Officer and the Network Security Administrator, by function. The HIPAA Compliance Officer is responsible for overseeing controlled access by other program staff to the HIPAA data. The Network Security Administrator is responsible for controlling access to the program servers by authorized users and locations determined by the HIPAA Compliance Officer.

Intellectual property rights

Entities or persons who will hold the intellectual property rights to the data, and how IP will be protected if necessary. Any copyright constraints (e.g., copyrighted data collection instruments) should be noted.

Generic Example 1: The principal investigators on the project and their institutions will hold the copyright for the research data they generate.

Generic Example 2: The principal investigators on the project and their institutions will hold the copyright for the research data they generate but will grant redistribution rights to [repository] for purposes of data sharing.

Generic Example 3: The data gathered will use a copyrighted instrument for some questions. A reproduction of the instrument will be provided to [repository] as documentation for the data deposited with the intention that the instrument be distributed under "fair use" to permit data sharing, but it may not be redisseminated by users.

Access and sharing

A description of how data will be shared, including access procedures, embargo periods, technical mechanisms for dissemination and whether access will be open or granted only to specific user groups. A timeframe for data sharing and publishing should also be provided.

Generic Example 1: The research data from this project will be deposited with [repository] to ensure that the research community has long-term access to the data.

Generic Example 2: The project team will create a dedicated Web site to manage and distribute the data because the audience for the data is small and has a tradition of interacting as a community. The site will be established using a content management system like Drupal or Joomla so that data users can participate in adding site content over time, making the site self-sustaining. The site will be available at a .org location. For preservation, we will supply periodic copies of the data to [repository]. That repository will be the ultimate home for the data.

Generic Example 3: The research data from this project will be deposited with [repository] to ensure that the research community has long-term access to the data. The data will be under embargo for one year while the investigators complete their analyses.

Generic Example 4: The research data from this project will be deposited with the institutional repository on the grantees' campus.

Audience

The potential secondary users of the data.

Generic Example 1: The data to be produced will be of interest to demographers studying family formation practices in early adulthood across different racial and ethnic groups. In addition to the research community, we expect these data will be used by practioners and policymakers.

Selection and retention periods: A description of how data will be selected for archiving, how long the data will be held, and plans for eventual transition or termination of the data collection in the future.

Generic Example 1: Our project will generate a large volume of data, some of which may not be appropriate for sharing since it involves a small sample that is not representative. The investigators will work with staff of the [repository] to determine what to archive and how long the deposited data should be retained.

Archiving and preservation

The procedures in place or envisioned for long-term archiving and preservation of the data, including succession plans for the data should the expected archiving entity go out of existence.

Generic Example 1: By depositing data with [repository], our project will ensure that the research data are migrated to new formats, platforms, and storage media as required by good practice.

Generic Example 2: In addition to distributing the data from a project Web site, future long-term use of the data will be ensured by placing a copy of the data into [repository], ensuring that best practices in digital preservation will safeguard the files.

Ethics and privacy

A discussion of how informed consent will be handled and how privacy will be protected, including any exceptional arrangements that might be needed to protect participant confidentiality, and other ethical issues that may arise.

Generic Example 1: For this project, informed consent statements will use language that will not prohibit the data from being shared with the research community.

Generic Example 2: The following language will be used in the informed consent: The information in this study will only be used in ways that will not reveal who you are. You will not be identified in any publication from this study or in any data files shared with other researchers. Your participation in this study is confidential. Federal or state laws may require us to show information to university or government officials [or sponsors], who are responsible for monitoring the safety of this study.

Generic Example 3: The proposed medical records research falls under the HIPAA Privacy Rule. Consequently, the investigators will provide documentation that an alteration or waiver of research participants' authorization for use/disclosure of information about them for research purposes has been approved by an IRB or a Privacy Board.

Budget

The costs of preparing data and documentation for archiving and how these costs will be paid.
Requests for funding may be included.

Generic Example 1: Staff time has been allocated in the proposed budget to cover the costs of preparing data and documentation for archiving. The [repository] has estimated their additional cost to archive the data is [insert dollar amount]. This fee appears in the budget for this application as well.

Data organization

How the data will be managed during the project, with information about version control, naming conventions, etc.

Generic Example 1: Data will be stored in a Concurrent Versions System (CVS) system and checked in and out for purposes of versioning. Variables will use a standardized naming convention consisting of a prefix, root, and suffix system. Separate files will be managed for the two kinds of records produced: one file for respondents and another file for children with merging routines specified.

Quality Assurance

Procedures for ensuring data quality during the project.

Generic Example 1: Quality assurance measures will comply with the standards, guidelines, and procedures established by the World Health Organization.

Generic Example 2: For quantitative data files, the [repository] ensures that missing data codes are defined, that actual data values fall within the range of expected values and that the data are free from wild codes. Processed data files are reviewed by a supervisory staff member before release.

Legal requirements

A listing of all relevant federal or funder requirements for data management and data sharing.

Generic Example 1: The proposed medical records research falls under the HIPAA Privacy Rule. Consequently, the investigators will provide documentation that an alteration or waiver of research participants' authorization for use/disclosure of information about them for research purposes has been approved by an IRB or a Privacy Board.