Suggested Elements to Cover in a Data Management Plan

(Adapted from [ICPSR Guidelines for Effective Data Management Plans](#))

Elements that are HIGHLY RECOMMENDED

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.
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	A description of the metadata to be provided along with the generated data, and a discussion of the metadata standards used.
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.
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. Possible mechanisms for archiving and sharing include: 1) Domain repository like ICPSR (social science); 2) 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*; 3) 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; 4) 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.
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.

Elements considered OPTIONAL

ELEMENT	DESCRIPTION
Existing data	A survey of existing data relevant to the project and a discussion of whether and how these data will be integrated.

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.
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.
Budget	The costs of preparing data and documentation for archiving and how these costs will be paid. Requests for funding may be included.
Legal requirements	A listing of all relevant federal or funder requirements for data management and data sharing.
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.

Note: Individual Data Management Plans should be tailored to the specific research activities described within each proposal.

Data Management Plan [TEMPLATE with Examples]

1. **Data description** (Highly Recommended)

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.

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.

2. **Format** (Highly Recommended)

Example 1: Quantitative survey data files generated will be processed and submitted to the [repository] as SPSS system files with 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.

Example 2: Digital video data files generated will be processed and submitted to the [repository] in MPEG-4 (.mp4) format.

Example 3: Digital image data will be processed and submitted to the [repository] in TIFF version 6 uncompressed (.tif) format.

Example 4: Geospatial data will be processed and submitted to the [repository] as an ESRI Shapefile (essential - .shp, .shx, .dbf, optional - .prj, .sbx, .sbn).

Example 5: Textual data will be processed and submitted to the [repository] as plain text data, ASCII (.txt).

3. **Metadata** (Highly Recommended)

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.

Example 2: The clinical data collected from this project will be documented using CDISC metadata standards.

4. **Storage and backup** (Highly Recommended)

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.

5. **Intellectual property rights** (Highly Recommended)

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

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

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.

6. **Access and sharing** (Highly Recommended)

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.

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.

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.

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

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 5: The data generated by this project will not pose a disclosure risk. All data will be de-identified before posting to the Web site established by the principal investigators.

Example 6: This project will generate data linked to administrative records, so the data will be distributed through a restricted data use agreement managed by [repository]. Through this mechanism, users will apply to use these files, create data security plans, and agree to other access controls.

Example 7: Because the data generated will cover sensitive topics, it is expected that the data will be deposited with [repository] and distributed through the secure data enclave mechanism, requiring researchers to visit the enclave to access the data under secure conditions.

Indicate when the data will be made available to others.

Example 8: The research data from this project will be deposited with [repository] before the

end of the project so that any issues surrounding the usability of the data can be resolved.

Example 9: The data will be deposited with [repository] but not disseminated for one year to give the investigators time to publish their findings.

7. **Archiving and preservation** (Highly Recommended)

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.

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.

8. **Ethics and privacy** (Highly Recommended)

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

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.

Example 3: The following language will be used for video data: Participants in this study will be videotape recorded. The videos will be made available through the Web for others to use. However, all users will be required to use the videos for research purposes only and will not be allowed to share the information from the videos with others.

If applicable, what are your plans to obtain IRB approval?

Example 4: For this project, the principal investigators will request expedited IRB review compliant with procedures established by the [University] campus IRB. Research activities envisioned present no more than minimal risk to human subjects.

Example 5: Because the project involves more than minimal risk to human subjects, the project will undergo full IRB board review, as required by federal regulations.

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

Example 6: 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.

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

Example 7: During data analysis, the data will be accessible only by certified members of the project team. The research project will remove any direct identifiers in the data before deposit

with [repository].

9. **Existing data** (Optional)

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...

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.

10. **Data organization** (Optional)

Example 1: Data will be stored in a CVS system and checked in and out for purposes of versioning. Variables will use a standardized naming convention consisting of a prefix, root, 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.

11. **Quality assurance** (Optional)

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

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.

12. **Security** (Optional)

Example 1: The data will be processed and managed in a secure non-networked environment using virtual desktop technology.

Example 2: The data files from this study will be managed, processed, and stored in a secure environment (e.g., lockable computer systems with passwords, firewall system in place, power surge protection, virus/malicious intruder protection) and by controlling access to digital files with encryption and/or password protection. Deidentifed files will be deposited with [repository] whose security policy has been written according to best practices.

13. **Responsibility** (Optional)

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.

Example 2: All research data collected as part of this project is owned by the University. The Principal Investigator of this project will take responsibility for the collection, management, and sharing of the research data.

14. **Budget** (Optional)

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.

15. **Legal requirements** (Optional)

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.

16. **Audience** (Optional)

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.

Example 2: In addition to the research community, we expect these data will be used by practioners and policymakers.

17. **Selection and retention periods** (Optional)

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.

Example 2: Our research project will generate data from a large national sample. These data will be retained by [repository] as part of their permanent collection.