Sample Language for DMPs

(Samples modified from Inter-University Consortium for Political and Social Science Research (ICPSR) webinar held January 2011 on "Preparing A Data Management Plan")

Data Description – [Insert a brief description of the information to be gathered -- the nature, scope, and scale of the data that will be generated or collected.]

> *Sample Language:*   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 release through a restricted use contract. These data, which will be submitted to the Inter-University Consortium for Political and Social Science Research (ICPSR) archives, fit within the scope of the ICPSR Collection Development Policy. A letter of support describing ICPSR's commitment to the data as they have been described is provided.

Access and Sharing – [Insert 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.

> *Sample Language #1:* The research data from this project will be deposited with {insert repository name} to ensure that the research community has long-term access to the data.

> *Sample Language # 2:* The research data from this project will be deposited with the digital repository of the Inter-university Consortium for Political and Social Research (ICPSR) to ensure that the research community has long-term access to the data. The integrated data management plan proposed leverages capabilities of ICPSR and its trained archival staff. ICPSR will make the research data from this project available to the broader social science research community. *Public-use data files*: These files, in which direct and indirect identifiers have been removed to minimize disclosure risk, may be accessed directly through the ICPSR Web site. After agreeing to Terms of Use, users with an ICPSR MyData account and an authorized IP address from a member institution may download the data, and non-members may purchase the files. *Restricted-use data files*: These files are distributed in those cases when removing potentially identifying information would significantly impair the analytic potential of the data. Users (and their institutions) must apply for these files, create data security plans, and agree to other access controls. *Timeliness*: The research data from this project will be supplied to ICPSR before the end of the project so that any issues surrounding the usability of the data can be resolved. Delayed dissemination may be possible. The Delayed Dissemination Policy allows for data to be deposited but not disseminated for an agreed-upon period of time (typically one year).

Metadata – [Insert a description of the metadata to be provided along with the generated data, and a discussion of the metadata standards used.]

> *Sample Language #1:* Metadata will be tagged in XML using the Digital 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.

Sample Language #2: Substantive metadata will be provided in compliance with the most relevant standard for the social, behavioral, and economic sciences -- the Data Documentation Initiative (DDI). This XML standard provides for the tagging of content, which facilitates preservation and enables flexibility in display. These types of metadata will be produced and archived:

- *Study-Level Metadata Record*. A summary DDI-based record will be created for inclusion in the searchable ICPSR online catalog. This record will be indexed with terms from the ICPSR Thesaurus to enhance data discovery.
- *Data Citation with Digital Object Identifier (DOI)*. A standard citation will be provided to facilitate attribution. The DOI provides permanent identification for the data and ensures that they will always be found at the URL specified.
- *Variable-Level Documentation*. ICPSR will tag variable-level information in DDI format for inclusion in ICPSR's Social Science Variables Database (SSVD), which allows users to identify relevant variables and studies of interest.
- *Technical Documentation*. The variable-level files described above will serve as the foundation for the technical documentation or codebook that ICPSR will prepare and deliver.
- *Related Publications*. Resources permitting, ICPSR will periodically search for publications based on the data and provide two-way linkages between data and publications.

Intellectual Property Rights – [Indicate the 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. Refer to the University Policy Manual for more on Copyrights and Patents.]

Sample Language #1: The PI(s) and their institution(s) will hold the copyright for the research data they generate.

Sample Language #2: The PI(s) and their institution(s) hold the copyright for the research data they generate. By depositing with ICPSR, investigators do not transfer copyright but instead grant permission for ICPSR to redisseminate the data and to transform the data as necessary to protect respondent confidentiality, improve usefulness, and facilitate preservation.

Ethics and Privacy – [Discuss 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.]

Sample Language #1: The informed consent statements for this project will use language that will not prohibit the data from being shared with the research community.

Sample Language # 2(when disclosing risk management): The following language will be used in the informed consent statement: 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 format his study or in any data files shared with other researchers.

Format – [Insert a description of the formats in which the data will be generated, maintained, and made available, including a justification for the procedural and archival appropriateness of those formats.]

> *Sample Language #1: Digital video data files generated will be processes and submitted to the {insert repository name} in MPEG-4 format.*

> *Sample Language #2:* The data and documentation will be submitted to ICPSR in recommended formats. ICPSR will make the quantitative data files available 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 in accordance with prevailing standards and practice. Currently, ICPSR stores quantitative data as ASCII along with setup files for the statistical software packages, and documentation is preserved using XML and PDF/A.

Archiving and Preservation – [Describe 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.]

> *Sample Language #1:* By depositing our data with {insert name of repository}, our project will ensure that the research data are migrated to new formats, platforms, and storage media as required by the current practice.

> *Sample Language #2:* ICPSR is a data archive with a nearly 50-year track record for preserving and making data available over several generational shifts in technology. ICPSR will accept responsibility for long-term preservation of the research data upon receipt of a signed deposit form. This responsibility includes a commitment to manage successive iterations of the data if new waves or versions are deposited. ICPSR will ensure that the research data are migrated to new formats, platforms, and storage media as required by good practice in the digital preservation community. Good practice for digital preservation requires that an organization address succession planning for digital assets. ICPSR has a commitment to designate a successor in the unlikely event that such a need arises.

Storage and Backup – [Describe 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.]

> *Sample Language #1:* [Insert name of repository} will place a master copy of each file (i.e., research data files, documentation, and other related files) in archival storage, with several copies store at designated locations and synchronized with the master through the storage resource broker.

> *Sample Language #2:* Research has shown that multiple locally and geographically distributed copies of digital files are required to keep information safe. Accordingly, ICPSR will place a master copy of each digital file (i.e., research data files, documentation, and other related files) in ICPSR's Archival Storage, with several copies stored with partner organizations at designated locations and synchronized with the master.

Additional Topics to address could include:

- A survey of existing data relevant to the project
- Management of data during the project including version control and naming convention.
- Procedures for ensuring data quality
- Technical and procedural protections for information.
- Cost of preparing data and documentation for archiving.
- Legal requirements
- Audience