ICPSR

Guidelines for Effective Data Management Plans

Data Management Plans

Federal funding agencies are increasingly recommending or requiring formal data management plans with all grant applications.

To help researchers meet those requirements, ICPSR offers these guidelines.

Based on our Data Management Plan Web site, this document contains a framework, example data management plans, links to other resources, and a bibliography of related publications.

ICPSR also hosts a blog on data management plans, and a recent webinar on the subject can be viewed on our Web site.

We hope you find this information helpful as you craft a data management plan. Please contact us at netmail@icpsr.umich.edu with any comments or suggestions.

Table of Contents

Framework for Creating a Data Management Plan

This framework takes the form of a list of elements with issues and questions to consider and examples for each element. Also included is information on why each element is important and suggestions for additional reading.

Data Description (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 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.

Access and Sharing (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 ICPSR (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 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 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.

Metadata (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 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.

Intellectual Property Rights (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 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.

Ethics and Privacy (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 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].

Format (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.

Preservation formats should be platform-independent and non-proprietary to ensure that they will be usable in the future.

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).

Archiving and Preservation (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. Preservation of digital information is widely considered to require more constant and ongoing attention than preservation of other media. Depositing data resources with a trusted digital archive can ensure that they are curated and handled according to good practices in digital

preservation.

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.

Storage and Backup (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 data are fragile and best practice for protecting them is to store multiple copies in multiple locations.

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 (Recommended)

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 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.

Responsibility (Recommended)

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 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.

Existing Data (Recommended)

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 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.

Selection and Retention Periods (Recommended)

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
Not all data need to be preserved in perpetuity, so thinking through the proper retention period for the data is important, in particular when there are reasons the data will not be preserved permanently.

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.

Audience (Recommended)

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 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.

Data Organization

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 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.

Quality Assurance

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 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.

Budget

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 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.

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 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.

Data Management Plan Resources and Examples

Resources for Development

- Australian National Data Service - [Data management planning](#)

- Digital Curation Centre - [Data Management Plans](#)

- Geoscience Australia - [Guide to Preparation of Data Management Plans](#)

- New South Wales Natural Resources Information Management Strategy - [Data Management Planning Guidelines](#)

- University of Oxford - [Data Management Planning Checklist](#)

- University of Oxford - [Data Management Planning](#)

Templates & Tools

- Australia National University - [Information Literacy Program DPM Template](#)

- caBIG Data Sharing and Intellectual Capital Knowledge Center - [caBIG Data Sharing and Security Framework (DSSF)](#)

- Digital Curation Centre - [Data Management Plan Content Checklist](#)

- Digital Curation Centre - [DMP Online Data Management Planning Tool](#)

- Queensland University of Technology (QUT) - [Data Management Checklist](#)

- UK Rural Economy and Land Use Programme - [Data Management Plan](#)

- University of Melbourne - [Research Data Management Plan Template](#)

Guidance on Funder Requirements

- Association of Research Libraries - [Guide for Research Libraries: The NSF Data Sharing Policy](#)

- Digital Curation Centre - [Summary of UK research funders' expectations for the content of data management and sharing plans](#)

- University of Minnesota Libraries - [Funding Agency and Data Management Guidelines](#)

- University of Oxford - [Research Data Management Funder Policies](#)

University Data Management Web Sites

- Cornell University - Research Data Management Service Group

- MIT Libraries - Manage Your Data: Data Management: Subject Guides

- University of Colorado, Colorado Clinical and Translational Sciences Institute - Data Management Best Practices

- University of Edinburgh - Research data management guidance

- University of Melbourne - Research Data Management for Researchers

- University of Minnesota Libraries - Funding Agency and Data Management Guidelines

- University of Oxford - Research Data Management Website

Good Practice Guidance

- Australia National University - ANU Data Management Manual, September 2010

- Charles Beagrie Limited - User Guide for Keeping Research Data Safe

- Columbia University, Center for International Earth Science Information Network (CIESIN) - Guide to Managing Geospatial Electronic Records, June 2005

- Council of European Social Science Data Archives (CESSDA) - Sharing Data Website

- Finnish Social Science Data Archive (FSSDA) - Data Management Planning Website

- Gary King - Data Sharing and Replication Information

- Inter-university Consortium for Political and Social Research (ICPSR) - Guide to Social Science Data Preparation and Archiving, 2009

- Mark Mosley, Data Management Association - DAMA-DMBOK Functional Framework Version 3.02, September 10, 2008

- National Academy of Sciences - Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age

- National Science Board (NSB) - Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century

- New South Wales Natural Resources Information Management Strategy - Guiding Principles for Data Management

- UK Data Archive - Managing and Sharing Data: A Best Practice Guide for Researchers, September 2009

- UK Data Archive - [Data Management Practices in the Social Sciences, August 10, 2010](#)

- UK Data Archive - [Create and Manage Data Website](#)

Federal Agency Policies on Data Management and Sharing

- Centers for Disease Control and Prevention (CDC) - [Policy on Releasing and Sharing Data](#)

- Council on Governmental Relations - [Access to and Retention of Research Data: Rights and Responsibilities, March 2006.](#)

- National Institute of Justice - [Data Resources Program](#)

- National Institutes of Health - [Data Sharing Policy](#)

- National Science Foundation - [Proposal and Award Policies and Procedures Guide](#)

- National Science Foundation, Social, Behavioral and Economic Sciences (SBE) - [Data Management Plan Policy](#)

- National Science Foundation, Social, Behavioral and Economic Sciences (SBE) - [Data Archiving Policy](#)

- National Science Foundation (NSF) - [Dissemination and Sharing of Research Results](#)

- National Science Foundation Directorate for Engineering - [Data Management for NSF Engineering Directorate Proposals and Awards](#)

- Office of Management and Budget (OMB) - [Circular A-110, September 30, 1999](#)

Other Data Management Plan Examples from Natural Sciences

- [Atmospheric Radiation Measurement Program (ARM)](#)

- Carbon Dioxide Information Analysis Center, Environmental Sciences Division, Oak Ridge National Laboratory, U.S. Dept. of Energy - [WHPO/CCHDO Data Management Plan for CTD/Hydrographic/CO](#)

- Community Climate System Model (CCSM) - [Data Management Plan](#)

- Environmental Protection Agency - [Deepwater Horizon Rig Explosion Air Sampling & Monitoring Data Management Plan](#)

- Geospatial Data Preservation - [Geospatial Data Preservation Resource Center](#)

- NASA - [Mission Project Data Management Plans Website](#)

- National Institutes of Health Division of AIDS (DAIDS) - [Data Management Plan Template for laboratory studies](#)

- North American Monsoon Experiment (NAME) - [Data Management Plan, Draft version 1.0, August 2004](#)

- North American Regional Climate Change Assessment Program - [Operational Data Management Plan, 9/28/07](#)

- Plate Boundary Observatory - [Data Management Plan](#)

- S. Jeffers, Southwest Research Institute - [Project Data Management Plan for the Mars Express ASPERA-3 Processing and Archiving Facility, July 10, 2003](#)

- South American Low-Level Jet Experiment (SALLJEX) - [Data Management Plan, December 2002](#)

- Southwest Research Institute - [IBEX Project Data Management Plan, July 2007](#)

- Space Science and Engineering at the Southwest Research Institute - [New Horizons Data Management and Archiving Plan](#)

- Tennessee Valley Authority Kingston Ash Recovery Project - [Data Management Plan](#)

- U.S. Array - [Data Management Plan, 8/2/04](#)

- U.S. Department of the Interior, National Parks Service Inventory and Monitoring Program - [Data Management Plan for the Northern Great Plains Inventory and Monitoring Network, March 2009](#)

- VORTEX2 - [Data Management Policy, 4/15/09](#)

- WMO-IOC Joint Technical Commission for Oceanography and Marine Meteorology - Data Management Plan Implementation Details April 2009

- WMO-IOC Joint Technical Commission for Oceanography and Marine Meteorology - Data Management Plan 2008

- WMO-IOC Joint Technical Commission for Oceanography and Marine Meteorology - Data Management Plan Website

Appendix A: Elements of a Data Management Plan

Element	Description	Recommended?	NSF Mapping
Data description	A description of the information to be gathered; the nature and scale of the data that will be generated or collected.	Yes	Expected Data
Existing data	A survey of existing data relevant to the project and a discussion of whether and how these data will be integrated.	Yes	Expected Data
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.	Yes	Data Format and Dissemination
Metadata	A description of the metadata to be provided along with the generated data, and a discussion of the metadata standards used.	Yes	Data Format and Dissemination
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.	Yes	Data Storage and Preservation of Access
Security	A description of technical and procedural protections for information, including confidential information, and how permissions, restrictions, and embargoes will be enforced.	Yes	Data Format and Dissemination
Responsibility	Names of the individuals responsible for data management in the research project.	Yes	Roles and Responsibilities
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.	Yes	Data Format and Dissemination

Element	Description	Recommended?	NSF Mapping
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.	Yes	Data Storage and Preservation of Access
Audience	The potential secondary users of the data.	Yes	Data Format and Dissemination
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.	Yes	Data Format and Dissemination
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.	Yes	Data Storage and Preservation of Access
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.	Yes	Data Format and Dissemination
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.		

Appendix B: Sample Data Management Plan for Depositing Data with ICPSR

This sample plan is provided to assist grant applicants in creating the required Data Management Plans. Researchers should feel free to edit and customize this text before submission. Note that a letter of commitment from ICPSR confirming that it will archive the data should accompany the plan. Please contact ICPSR Director of Acquisitions Amy Pienta, apienta@umich.edu, to request such a letter.

Please review our guidelines on how to deposit data with ICPSR.

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

Designated Archive – 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.

Access and Sharing – 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 – 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 – Principal investigators and their institutions 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 – *Informed consent*: For this project, informed consent statements, if applicable, will not include language that would prohibit the data from being shared with the research community.

Disclosure risk management: The research project will remove any direct identifiers in the data before deposit with ICPSR. Once deposited, the data will undergo procedures to protect the confidentiality of individuals whose personal information may be part of archived data. These include: (1) rigorous review to assess disclosure risk, (2) modifying data if necessary to protect confidentiality, (3) limiting access to datasets in which risk of disclosure remains high, and (4) consultation with data producers to manage disclosure risk. ICPSR will assign a qualified data manager certified in disclosure risk management to act as steward for the data while they are being processed. The data will be processed and managed in a secure non-networked environment using virtual desktop technology.

Format – *Submission*: The data and documentation will be submitted to ICPSR in [recommended formats](#).

Access: 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.

Preservation: 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 – 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 – 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.

Appendix C: Further Reading on Data Management Issues

Below are suggestions for additional readings germane to the elements of a data management plan.

Abrahamson, M., K. Bollen, M. Gutmann, G. King, and A. M Pienta. 2009. Preserving Quantitative Research-Elicited Data for Longitudinal Analysis: New Developments in Archiving Survey Data in the U.S. *Historical Social Research* 34(3), 51-59.

Beagrie, Charles. "User Guide for Keeping Research Data Safe: Assessing Costs/Benefits of Research Data Management, Preservation and Re-use." Version 1.0, December 2010. See especially the Benefits Taxonomy, pages 24-30.

Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age, National Academy of Sciences. Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age. Washington, DC: National Academy Press, 2009. http://www.nap.edu/catalog.php?record_id=12615

Green, Ann, and Myron P. Gutmann. "Building Partnerships Among Social Science Researchers, Institution-based Repositories and Domain Specific Data Archives." http://deepblue.lib.umich.edu/handle/2027.42/41214

Gregory, Arofan, Pascal Heus, and Jostein Ryssevik. "Metadata." Working Paper No. 57. German Council for Social and Economic Data (RatSWD). March 2009. http://www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_57.pdf

Gutmann, M.P., K. Schürer, D. Donakowski, and H. Beedham. 2004. "The Selection, Appraisal, and Retention of Social Science Data." *Data Science Journal* 3:209-221. http://www.jstage.jst.go.jp/article/dsj/3/0/209/_pdf

Gutmann, M.P., M. Abrahamson, M. O. Adams, M. Altman, C. Arms, K. Bollen, M. Carlson, J. Crabtree, D. Donakowski, G. King, J. Lyle, M. Maynard, A. Pienta, R. Rockwell, L. Timms-Ferrara, and C. Young. 2009. "From Preserving the Past to Preserving the Future: The Data-PASS Project and the Challenges of Preserving Digital Social Science Data." *Library Trends* 57: 315-337.

Gutmann, M.P., K. Witkowski, C. Colyer, J.M. O'Rourke, and J. McNally. 2008. "Providing Spatial Data for Secondary Analysis. Issues and Current Practices relating to Confidentiality." *Population Research and Policy Review* 27:639-665. http://dx.doi.org/10.1007/s11113-008-9095-4.

Inter-university Consortium for Political and Social Research (ICPSR). ICPSR Guide to Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle. 5th Edition, 2012.
http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf

King, Gary. "An Introduction to the Dataverse Network as an Infrastructure for Data Sharing," [with discussion] *Sociological Methods and Research*, Vol. 32, No. 2 (November,

2007): Pp. 173-199.

King, Gary. "Replication, Replication," PS: Political Science and Politics, with comments from nineteen authors and a response, "A Revised Proposal, Proposal," Vol. XXVIII, No. 3 (September, 1995): Pp. 443-499.

LeClere, Felica. 2010. Too Many Researchers Are Reluctant to Share Their Data. Chronicle of Higher Education, August 3, 2010.

National Science Board. Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century, December 2005.
http://www.nsf.gov/pubs/2005/nsb0540/

OECD Principles and Guidelines for Access to Research Data from Public Funding.
http://www.oecd.org/dataoecd/9/61/38500813.pdf

O'Rourke, JM, Roehrig, S, Heeringa, SG, Reed, BG, Birdsall, WC, Overcashier, M., Zidar, K. "Solving Problems of Disclosure Risk While Retaining Key Analytic Uses of Publicly Released Microdata." *Journal of Empirical Research on Human Research Ethics*, 1(3), September. 2006, 63-84.

Pienta, Amy Mehraban, JoAnne McFarland O'Rourke, Melissa M. Franks. Forthcoming. "Getting Started: An Introduction to Working with Archival Data Sets." In Kali H. Trzesniewski, M. Brent Donnellan, & Richard E. Lucas (Eds.) *Analyzing Archival Data: Methods and Illustrations*.

Pienta, Amy M., George Alter, and Lared Lyle. 2010. "The Enduring Value of Social Science Research: The Use and Reuse of Primary Research Data." Presented at the BRICK, DIME, STRIKE Workshop, The Organisation, Economics, and Policy of Scientific Research, Turin, Italy, April 23-24, 2010
http://hdl.handle.net/2027.42/78307

Van Wey, L., R. Rindfuss, M. P. Gutmann, B. Entwisle, and D. Balk. 2005. "Confidentiality and Spatially Explicit Data: Concerns and Challenges." Proceedings of the National Academy of Science 102(43): 15337-15342.

Vardigan, Mary , Pascal Heus, Wendy Thomas. "Data Documentation Initiative: Toward a Standard for the Social Sciences." *International Journal of Digital Curation*, Vol 3, No 1 (2008).
http://www.ijdc.net/index.php/ijdc/article/view/66