Elements of Data Management Plans: A Gap Analysis and Recommendations

Linda Detterman, Peter Granda, Amy Pienta, and Mary Vardigan
ICPSR

IASSIST 2011 – Vancouver
Friday, June 3

ICPSR | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

Today's presentation

- NSF requirement
- ICPSR's response
- Effects of NSF requirement
- Challenging issues

The statement that rocked the research world

- The National Science Foundation has released a new requirement for proposal submissions regarding the management of data generated using NSF support. Starting in January, 2011, all proposals must include a data management plan (DMP).
- The plan should be short, no more than two pages, and will be submitted as a supplementary document. The plan will need to address two main topics:
  - *What data are generated by your research?*
  - *What is your plan for managing the data?*

Specifically from NSF SBE

Quantitative Social and Economic Data Sets
For appropriate data sets, researchers should be prepared to place their data in fully cleaned and documented form in a data archive or library within one year after the expiration of an award. Before an award is made, investigators will be asked to specify in writing where they plan to deposit their data set(s). This may be the *Inter-University Consortium for Political and Social Research (ICPSR) at the University of Michigan*, but other public archives are also available. The investigator should consult with the program officer about the most appropriate archive for any particular data set.

ICPSR INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

In response

ICPSR decided to

- Investigate what others had been doing in this area -- specifically, determine the most commonly used elements of data management plans
- Create a set of Web pages
- Provide assistance to researchers in preparing plans

Environmental scan of guidance

- Australia National University Information Literacy Template
- Australian National Data Service Data Management Planning document
- Digital Curation Centre DMP Checklist and Template
- Finnish Social Science Data Archive's Data Management Planning Website
- Geoscience Australia's Guide to Preparation of DMPs
- MIT Libraries' Data Management Web page
- National Science Board's *Long-Lived Digital Data Collections*
- NSF Directorate for Engineering's Data Management Guidance
- Queensland University of Technology Data Management Checklist
- UK Rural Economy and Land Use (RELU) Data Management Plan
- University of Melbourne's Research Data Management Plan Template

Elements matrix

ICPSR | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

Deposit Data & Findings
How to Deposit
Data Deposit Form
Data Preparation Guide
Data Management Plans
Preserving Respondent Privacy/Confidentiality
Suggest Data to Archive

Usage of Elements Across a Sample of Organizations

	ANU	ANDS	DCC	FSSDA	Geo Aus	MIT	NSB	NSF Eng	QUT	RELU	Melbourne
Data description	•	•	•	•	•	•	•	•	•	•	•
Existing data	•	•	•						•	•	
Format	•	•	•	•	•	•	•	•	•	•	•
Metadata		•	•	•			•	•	•		•
Data organization	•	•	•						•		•
Quality Assurance			•	•		•			•	•	•
Storage and backup	•	•	•	•	•	•			•	•	•
Security	•	•	•	•	•	•		•		•	
Responsibility	•	•	•		•	•		•	•	•	•
Budget	•	•	•					•			
IP	•	•	•	•		•		•	•	•	•
Legal requirements	•	•	•								
Access and sharing	•	•	•	•	•	•	•	•	•		•
Audience	•	•	•			•				•	
Selection and retention periods	•	•	•			•	•	•	•		•
Archiving and preservation		•	•	•		•	•	•			•
Ethics and privacy			•	•		•		•	•		•

ICPSR

NSF elements

- Roles and responsibilities
- Expected data
- Period of data retention
- Data formats and dissemination
- Data storage and preservation of access

Element set with mapping

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 Responsibility
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
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	Period of Data Retention
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

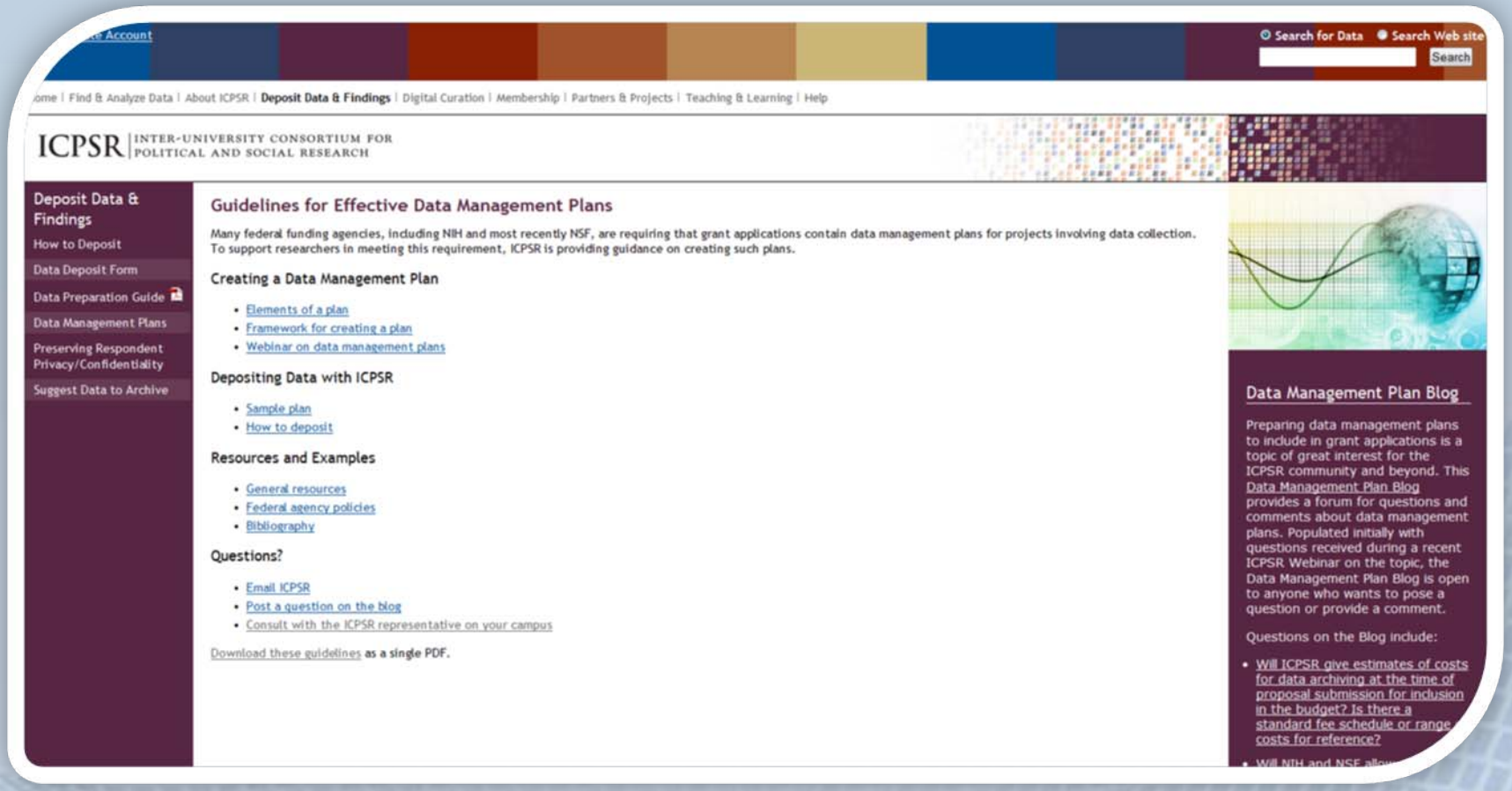
ICPSR webinars

November 2010 -- 134 attended

January 2011 -- 535 attended

February 2011 -- 71 attended

ICPSR's DMP site

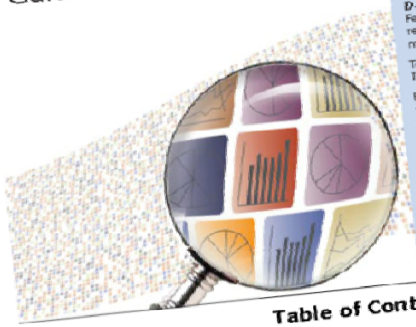


www.icpsr.umich.edu/ICPSR/dmp/

Downloadable guidelines

ICPSR | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH
A PARTNER IN SOCIAL SCIENCE RESEARCH

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 2
Data Management Plan Resources and Examples 11
Resources for Development 11
Templates and Tools 11
Guidance on Funder Requirements 12
University Data Management Web Sites 12
Good Practice Guidance 13
Federal Agency Policies on Data Management and Sharing 14
Other Data Management Plan Examples from Natural Sciences 16
Appendix A: Elements of a Data Management Plan 18
Appendix B: Sample ICPSR Data Management Plan 21
Appendix C: Data Management Bibliography 21

ICPSR | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

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

ICPSR | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

Framework for creating plans

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.

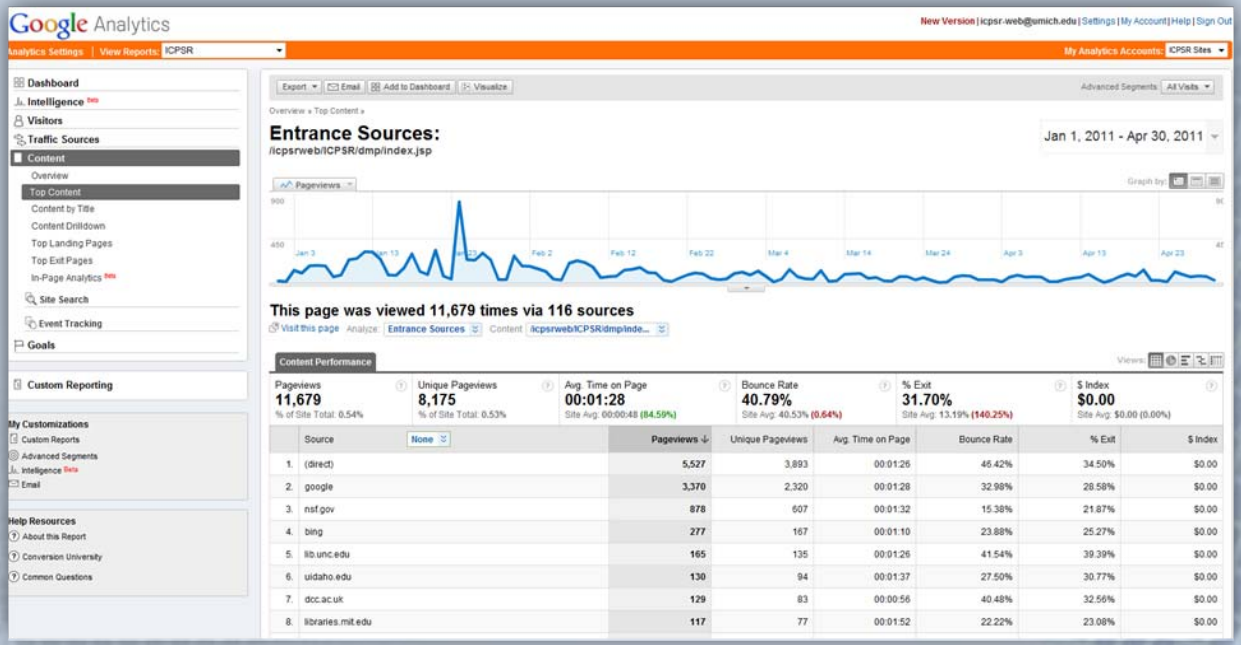
ICPSR's DMP blog

http://datamanagementplans.blogspot.com/

ICPSR | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

Web traffic – DMP page

- January 2011: 3,984 views
- January – April 2011: 7,802 views
- Where are they coming from?
  - 5,527 Direct (bookmarked, etc.)
  - 3,370 from Google search
  - 878 from NSF



ICPSR | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

Effects of the requirement

- ICPSR has received requests for letters of support, generally outside the traditional scope
- ICPSR has been involved in University of Michigan-wide meetings focused on data management and the University has slowly begun to establish its own set of guidelines for researchers about data sharing
- ICPSR became a contributing partner in the DMP Online project, working with UVa on social science

Challenging issues

- How will plans be assessed? How will the requirement be enforced?
- How might we attach costs to data curation and preservation activities? Should we?
- Some disciplines have no natural archive for their data, and institutional repositories cannot always meet these needs
- How does the role of data archives and service providers change in this new environment? Are new products necessary? What about self archiving?

Questions?

- Linda Detterman lindamd@umich.edu
- Peter Granda peterg@umich.edu
- Amy Pienta apienta@umich.edu
- Mary Vardigan vardigan@umich.edu