Data Preservation Alliance for the Social Sciences

(Data-PASS)

Final Report to NDIIPP at the Library of Congress

ICPSR
University of Michigan

IQSS and the Murray Archive
Harvard University

The Odum Institute
University of North Carolina-Chapel Hill

The Roper Center
University of Connecticut

and

The Electronic and Special Media Records Service Division
National Archives and Records Administration (NARA)

Table of Contents

Overview 3

Successes 4

Lessons Learned 7

Future Plans 9

Appendices

A Data-PASS Memorandum of Understanding 11

B List of Best Practices, Presentations, and Publications 16

Overview

The Data Preservation Alliance for the Social Sciences (Data-PASS) is a voluntary partnership of organizations created to archive, catalog and preserve data used for social science research. The partner organizations include the Inter-university Consortium for Political and Social Research, The Roper Center for Public Opinion Research, the Howard W. Odum Institute for Research in Social Science, the Electronic and Special Media Records Service Division of the National Archives and Records Administration, and the Henry A. Murray Research Archive.

Data-PASS came into existence in 2004 with funding from the Library of Congress' National Digital Information Infrastructure and Preservation Program (NDIIPP) and cost-share contributions from all partner institutions. The initial project was three years and was refunded and extended up through 2010. The goal of the initial project was to create a sustainable partnership model for identifying, appraising, acquiring, cataloging, and preserving "at risk" social science data, to “ensure the long-term preservation of the vital heritage of digital material that allows our nation to understand itself, its social organization, and its policies and politics through social science research.”

Throughout the years, the partners have made significant progress – drafting archival standards for the partnership, cooperating on various technology solutions for strengthening preservation and access, and rescuing data collections at risk of being lost. The Data-PASS partners have established working arrangements that have enhanced their individual organizations and improved preservation and access to digital information in their holdings.

To date, the partnership has identified thousands of at risk research studies and acquired over 1,300 data collections. These range from data collections created under National Science Foundation (NSF) and National Institutes of Health (NIH) grants, to surveys conducted by private research organizations, to state-level polling data, to data records created by governmental research or administrative programs. The partners have also established a shared electronic catalog for the tens of thousands of studies or series that comprise each partner’s entire data holdings. The Data-PASS shared catalog creates, for the first time ever, a unified way to find social science data in major archives.

A founding member of the National Digital Stewardship Alliance, the Data-PASS partnership continues to archive and catalog social science data, extend technical research into replicated preservation of archived collections, and advocate best practices in digital preservation. One major ongoing initiative by the partners is the development of the *SAFE-Archive* system, an open-source auditing and provisioning platform for verified, policy-driven replication of digital holdings. The prototype system built under NDIIPP funds is being developed into a self-contained system that will allow any library, museum, or archive to easily replicate their content, and to audit the replicas of that content using an existing public or private LOCKSS network.

Successes

The six year Data-PASS partnership has resulted in numerous successful outcomes, particularly with respect to organizational teamwork, collection development, and technological improvements.

Collaborative Partnership – The Data-PASS partners have enjoyed a productive and friendly collaborative relationship, including jointly developed agreements, an open and interoperable catalog, and harmonized acquisition/processing processes. Each partner has brought one or more areas of expertise to the table, ready to share them with the other partners: Sometimes it is expertise in building software, sometimes it is expertise in managing a certain process, and other times it is establishing a protocol or standard for communication. Members of the partnership have had the opportunity to examine each of these shared capabilities for applicability to their own organizations.

For example, the Odum Institute worked closely with Harvard to test the Dataverse Network for data delivery and to suggest and plan enhancements. ICPSR shared its data processing and quality control technology with the Roper Center, which allows each organization to share the burden of processing polling data while enjoying the full benefits of the processed datasets. Harvard re-engineered their entire acquisition process to incorporate Data-PASS identification, acquisition, and best practice development. Collaboration continues as the Data-PASS partners grow the shared catalog, reach out to new partners, jointly advocate for citation standards, build and deploy a LOCKSS-based replicated storage platform.

Learning from the Digital Preservation Community – Participating in the NDIIPP program, including presenting and publishing, has brought the Data-PASS partners into a much larger world of digital content delivery and digital preservation, which has had invaluable benefits. The partners now feel part of the digital library community and have integrated key digital preservation research and examples into their workflows. For example, the partners have built in support for TRAC, the Trustworthy Repositories Audit & Certification Criteria and Checklist, as a critical component of the *SAFE-Archive* audit system that currently is under development. Moreover, the partnership aims to have the Data-PASS organization as a whole constitute a "trusted repository" under the TRAC criteria.

Shared Catalog of Social Science Data – Membership in Data-PASS has increased the visibility of member organizations and their collections. One key factor in this has been the unified catalog of collections operated by Harvard, which makes data easier to discover, and provides opportunities to build "virtual" collections that are more complete and comprehensive. This catalog completes the unification of social science data that has been a major goal of data archivists since the first Council of Social Science Data Archives in the 1960s.

Data Identified and Rescued – The partnership has identified thousands of potential data collections, with over 1,300 data collections archived. Some of the notable data archived include:

United States Information Agency (USIA) data – This work represents collaboration between the Roper Center for Public Opinion Research and the National Archives and Records Administration (NARA) to recover, preserve, document and make accessible public opinion survey data conducted on behalf of the United States Information Agency (USIA) Office of Research from 1952 through 1999. The USIA data collection consists of over 2,000 surveys conducted in dozens of countries that contributed to the formulation of US foreign and defense policy. Some of the surveys represent the only opinion surveys available from specific countries. The project enabled the archives to backfill and update each other's collections. Before, the data were hard to find and hard to use. Many of the data were preserved in NARA's textual holdings. Many of the datasets were in the obsolescent "multi-punched" format based on punched card technologies. Both the Roper Center and the U.S. National Archives now preserve and make available a much more complete set of the USIA Research Data Collection than either institution had pre-Data-PASS.

Longitudinal Study of Personality Development – Particularly notable among the rescued studies is the landmark Longitudinal Study of Personality Development, now archived at Harvard's Henry A. Murray Research Archive. This study, conducted by Jack and Jeanne Humphrey Block, is the most intensive study of human personality development in existence. Spanning a 30-year sweep of American history from the start of the Nixon Administration to the end of the 20th century and the dawn of the Digital Age, the study follows a sample of 128 children from age 3 through age 33 years, revisited on nine occasions. Assessments evaluated the domains of social, emotional, moral, cognitive, and ego development using more than 100 different psychological instruments measuring thousands of psychological variables. This data is also unique in linking an extraordinarily broad set of quantitative measures to qualitative transcripts and hundreds of gigabytes of videotaped interviews. This study has already been the subject of over one hundred publications, and contains a wealth of data that has yet to be analyzed, and can be expected to yield insights into human personality for decades to come.

Harris Poll Data – The Harris Poll is one of the major sources of social science related polls and was reputed to have more than six million panelists worldwide. The Odum Institute has been the archival home of the Lou Harris Data Center since the early 1960's. Over time, however, gaps had developed in the collection due to corporate transfers and personnel changes at Harris. During the initial phase of the Data-PASS project, Odum was able to reestablish contact with Harris Interactive and bring the entire series of datasets up to date. More than two years of data were recovered and archived. Odum continues to maintain this relationship to help serve the social science community an updated series of Harris Poll datasets in their entirety.

Philadelphia Social History Project (PSHP) – Through Data-PASS efforts, three extraordinary censuses (1838, 1847, and 1856) of the African-American population of Philadelphia, which were collected by Quaker and abolitionist organizations in response to attempts in the Pennsylvania legislature to disenfranchise non-whites, were rescued and archived. Among other unique types of information, these censuses report religious affiliation, wealth, and how those

born in slave states gained their freedom. The censuses were originally placed into machine-readable format by the Philadelphia Social History Project in the 1960s and 70s. The PSHP began as a comparative study of social mobility in a 19th-century city and grew into a broader example of "New Urban History." Studies of ethnic identities, family life, demography, and the use of urban space were based on a variety of sources, some of them unique.

Shared Stewardship of Content – Through the membership and management structure of the partnership, including data and intellectual rights as well as transfer protocols legal and technical policies and agreements, the Data-PASS partners enjoy shared stewardship of their collections. Replication and stewardship of content by multiple institutions constitutes a form of preservation insurance against funding loss; legal regime change; mission drift; uncontrolled curatorial modification; and institutional failure. Data-PASS is able to safeguard the collections of its members through transfer protocols, succession planning, and live replication of collections. If a member organization requires off-site replication of its collections, the partnership will provide it. And if a member organization is no longer institutionally capable of preserving and disseminating a collection, the collection can be preserved and disseminated through the partnership. Thus the Data-PASS partnership enhances each partner's core institutional commitments to preservation.

Syndicated Replication – The Data-PASS partnership successfully researched, deployed, and tested a prototype replicated storage platform for social science data. This prototype system was built around a core Private LOCKSS Networks (PLN) technology; a schema to encapsulate inter-archival replication commitments; an automated schema-driven service that audits (PLNs; and Open Archives (OAI-PMH) clients to harvest data collections from the Dataverse Network and other repositories using the Data Documentation Initiative (DDI) schema.

This prototype replication system supported collaboration among multiple institutions by supporting the auditing of resource commitments, and automatic harvesting of new collections. Each institution could commit to providing some amount of resources to a joint replication network. These commitments were expressed formally, and were auditable using the system. Moreover, when an institution wished to add a new collection to the preservation network, the system provided a way to automatically identify collaborating peers with the required resources, and to initiate regular harvesting by those peers.

Standard LOCKSS mechanisms and software were used for the harvesting framework (although like many others we have created our own plug-in), integrity checks, and recovery. The SSP also made use of a PLN tool, the "cache manager", which is a tool supplied by the LOCKSS group for monitoring PLN's. The cache manager was used to gather information on the state of the network.

The Data-PASS partners are now extending the prototype syndicated storage system into a self-contained *SAFE-Archive* platform that will allow any library, museum, or archive to easily replicate their content, and to audit the replicas of that content using an existing public or private LOCKSS network.

Lessons Learned

Data-PASS collaboration has allowed each partner to learn many lessons.

Importance of Metadata – When putting together the Data-PASS shared catalog, it became clear that metadata was to be the linchpin that supported many other services used by the partners, including resource discovery, resource identification and citation, resource location, resource administration, data integrity, provenance, access control, and layered services such as variable level search, reformatting, and on-line analysis. Given all of these services, the metadata that could be provided was very detailed, and required a social-science specific schema – Dublin Core, and other general purpose schemas were not going to be sufficient to meet all of the partnership needs. At the same time, metadata is expensive to produce. And once produced in a particular form, internal systems are built around it that make use of that particular form. Thus, to create the shared catalog IQSS needed to (1) set minimal requirements that did not impose a burden on partners, (2) gracefully add services as more metadata was provided, (3) provide tools to automatically extract and convert metadata from partner formats to the exchange formats and schemas used for the catalog. IQSS had anticipated initially that the metadata conversion process would be time consuming because of the need to build specialized tools. They were wrong – metadata conversion was time consuming because of the intellectual effort needed to understand each partners metadata and to map it to the exchange format at the abstract level. Once the intellectual heavy lifting was complete the tools already built for the Virtual Data Center, or which IQSS built for DVN worked quite well, and the technical effort to make such conversions was minimal.

Importance of the Early Data Life Cycle – Another lesson learned was the importance of building archiving into institutional processes. By seeking to archive data long after the data were collected, it is difficult to overcome such barriers as time, money, and knowledge for archiving. Giving social science data archives a role in helping to shape data collection promotes the early detection of data collection errors and problems of incomplete metadata. Archives have the ability to aid researchers in the collection of data to ensure not only the ease of analysis, but also the economical archiving of the end product.

ICPSR discovered these factors first hand during the project, when they focused their collection development activity toward federally sponsored social science research awards. They found that of the 6,565 awards they contacted, only about 20 percent of the responding projects had archived their data. Further, about half of the non-archived research was no longer available. Many principal investigators wrote about data that were either intentionally discarded or unintentionally lost. Even when data were still accessible, there were significant barriers to archiving. Several PIs indicated they shared a commitment to data sharing, but lacked the time and resources for preparing data for archiving.

Through Data-PASS, many of the partners have reshaped their collection development activities to emphasize early detection of and collaboration with data producers. ICPSR, for instance, has built a LEADS database of NSF and NIH awards likely to create social science data collections. Staff now proactively engage data creators early in the data life cycle, which is yielding positive results. Harvard

and the Odum Institute has embraced a new model of archiving, “virtual archiving,” and living collections that streamlines the archival process and captures data as they are produced, avoiding the need to “rescue” it later.

Teaming with Private Research Organizations – In the case of building relationships with Private Research Organizations (PROs), the partnership has had much success identifying, acquiring and processing social science data sets deemed to be ‘at-risk’ of being lost to future generations of researchers. Throughout the process, though, challenges materialized, and still remain, to promote a culture of data preservation and access among academic researchers, private research organizations, as well as government agencies. These challenges to relationship building generally fall into one or more of the four categories: attitude, economics, technology, and the broad category of logistics. While most researchers will agree about the need to preserve data in perpetuity, data acquisition attempts can be thwarted by the proprietary perspective of the principal investigator or by the level of substantive engagement of the private research firm. Additionally, all organizations and individual researchers are faced with economic realities and pressures to monetize their research outputs. The partner archives need to face the argument that if the data have value, then why should they be released to the archives without appropriate compensation? Technological barriers include the false sense of security that making data available on a web site is equivalent to data archiving, and the obsolescence of historical data formats and media, including punch cards. Finally, logistical obstacles persist at every point, including organizational policies, contractual obligations and the basic ability to locate storage media and appropriate documentation.

Efforts to establish and strengthen relationships with PROs have focused in part on earlier stages of the traditional archival process to include pre-embargoed release of survey data with restrictive agreements on the timing of redistribution. Additionally, partners have placed greater emphasis on evaluating the business interests and models employed by various PROs, in some cases identifying tools and service enhancements for use by the PROs in their own workflows. For example, the Roper Center currently provides search and data access services to Pew Research Center websites as a way to further strengthen the data contributor-data archive relationship. It is clear from this experience to work successfully with PROs the partners must recognize that each relationship is unique and requires a variety of approaches to encourage success. One of the primary approaches is to leverage existing services that PROs may not be in the position to build or maintain themselves, providing real, tangible benefits to participating in data preservation efforts.

Replicated Storage Lessons – While work to create and extend a replicated storage platform has proven feasibility, a lesson learned is that there are no "silver bullets" for solving the problem of archiving large collections of digital data. While LOCKSS appears to be a very promising technology and system for creating and maintain large-scale collections of digital content, it has also been clear that it would need extensive experimentation - and modification - for use with the Data-PASS collective of data archives. With generous funding from IMLS, and in collaboration with the LOCKSS team, we are extending the locks system successfully. The first production release of the SAFE-Archive software is planned for March, and is already being used internally.

Expanding the Partnership – Data-PASS expects to expand beyond the founding partners. In fact, Data-PASS has identified and contacted three new social science archives for possible inclusion in the Data-PASS partnership. Each of the prospective archives has expressed interest in joining, with one already committing to join.

Notwithstanding the interest, there remain legal, financial, and social hurdles to new partners joining the partnership. For instance, one of the prospective new partners is based at a university in a cash-strapped and politically unstable state, and the university's own policies toward computing, technology, and archiving are very decentralized and undeveloped. The archivist representing the prospective new partner is extremely excited by the potential to collaborate with colleagues and replicate the archive's at-risk holdings, but also hesitant to join Data-PASS until local policies are sorted out. Another prospective partner is located overseas, which brings into question which content legally can be replicated across international boundaries.

Future Plans

The Data-PASS partnership will continue after the contract with the Library of Congress has ended. In fact, the partners currently are in the process of signing an updated membership agreement. The ongoing partnership is voluntary, and is structured by a formal memorandum of understanding. Data-PASS partners commit to three things: Ratifying the membership agreement, which gives the partnership the licenses necessary to expose a member's content through the catalog, and to replicate it for preservation. Each partner also participates in virtual meetings.

There are no membership fees within Data-PASS. Partners contribute in-kind personnel time, and computing resources, which keeps overhead low. Partners also collaborate on grant proposals or other fundraising efforts to cover the costs of developing new tools and conducting research into archival practice.

The operational structure of the partnership continues to consist of two working groups, one called a "Steering Committee" composed of the leadership of each of the partner organizations, supplemented by others as needed. This group's regular teleconferences accomplish two goals, first to gauge the partnership's progress over time, and second to discuss future outcomes and set policy for the future. The second working group, the "Operations Committee," manages the week-to-week tasks, including the standards and procedures the partnering organizations use in the selection, appraisal, processing, and archiving and dissemination of the data.

All content that has been acquired throughout the NDIIPP project will continue to be distributed through the Data-PASS shared catalog after the contract with the Library of Congress has ended. Indeed, one of the primary tenets of the partnership is to replicate and preserve content among the partners, with transfer protocols in place should a partner institution no longer be able to retain the material they acquire. These policies help insure that partner data collections are protected against institutional and technological failure.

Data-PASS is interested in expanding beyond the founding partners, and is engaged in expanding the partnership to involve other organizations interested in preserving social science research.

Appendix A

Data-PASS: Data Preservation Alliance for the Social Sciences

Memorandum of Understanding

1. Mission

The Data Preservation Alliance for the Social Sciences (Data-PASS) is a voluntary partnership of organizations created to archive, catalog and preserve data used for social science research. The partnership grew out of the National Digital Information Infrastructure and Preservation Program (NDIIPP) established in 2004 by the U.S. Library of Congress for the preservation of historically significant digital content. This Memorandum of Understanding (MoU) sets forth the framework for both continuation of the work originated through the NDIIPP and the terms of membership in this partnership.

2. Objectives

The partnership emphasizes the following primary objectives:

a) Archive significant social science data collections.
b) Catalog and promote access to archived collections in the Data-PASS shared catalog.
c) Replicated preservation of archived collections.
d) Advocate best practices in digital preservation.

The partnership will endeavor to meet and uphold these objectives through the cooperation of and adherence to the terms set forth in this MoU, by its membership.

3. Benefits

Benefits of membership in the Data-PASS partnership include, but are not be limited to, the following:

3a. Expanded visibility of collections

Partner data collections are described on the Data-PASS shared catalog (http://dvn.iq.harvard.edu/dvn/dv/datapass), which harvests and exposes partner metadata to the partnership members and their constituent communities.

3b. Insurance against institutional and technological failure

Partner data collections are protected against institutional and technological failure through succession planning and replicated preservation.

3c. Advocacy for archival best practices

Partners leverage their substantial combined history and experience to advocate for and share archival best practices.

4. Membership

Based on the prior involvement in the NDIIPP, members of the Data-PASS partnership as of the time of the creation of the MoU include:

a) The Inter-university Consortium for Political and Social Research (ICPSR) at the University of Michigan
b) The Howard W. Odum Institute for Research in Social Science at the University of North Carolina at Chapel Hill
c) The Roper Center for Public Opinion Research at the University of Connecticut
d) The Harvard-MIT Data Center of the Institute for Quantitative Social Science at Harvard University
e) The Henry A. Murray Research Archive of the Institute for Quantitative Social Science at Harvard University
f) The custodial electronic records division of the National Archives and Records Administration (NARA)
g) The Library of Congress

Each partner carries equal weight when making decisions for the partnership. As set forth in detail in Section 6 below, this MoU presumes the flexibility to add members to the partnership.

5. Governance Structure

5a. Lead Partner

One partner, ICPSR, will be the lead organization for the partnership for a three-year fixed term, beginning January 1, 2011 and ending December 31, 2013. Prior to the completion of the three-year term, the Steering Committee (defined below) will determine who shall take the lead partner role thereafter. The lead partner will take responsibility for coordinating partnership efforts, including:

a) Scheduling conference calls.
b) Organizing meetings.
c) Retaining copies of all signed MoUs and Steering Committee notes and documentation (e.g., member applications, procedures, etc.)
d) Writing yearly reports highlighting partnership achievements.
e) Contributing personnel time to support the partnership objectives.

All other partners will support the lead organization, by:

a) Participating in conference calls and meetings.
b) Contributing personnel time to support the partnership objectives.

5b. Steering Committee

Major decisions for the partnership will be made by the Steering Committee, which is composed of at least one delegate from each of the partnering organizations. The Steering Committee will convene biannually by conference call. Each partner will carry equal weight when decisions are made.

5c. Operations Committee

In addition to the Steering Committee, an Operations Committee will hold quarterly conference calls to develop and coordinate the procedures and practices of the partnership, such as writing grant applications and appraising collections to be added to the Data-PASS shared catalog. This Committee will be made up of at least one operations staff member from each of the partners.

To meet the objectives of the partnership, representatives from each of the partners may also elect to meet in person once per year, in connection with professional conferences (e.g., IASSIST) to help defray costs associated with the Data-PASS partnership. The lead partner will facilitate communications related to enabling such meetings.

6. New Members

Other organizations may join the Data-PASS partnership. Applications for membership are made through the Steering Committee, which shall review qualifications and related requirements and determine acceptance for membership. New members must agree to meet the conditions of and sign this MoU.

7. Data and Intellectual Property Rights

The partnership will try to make all materials it collects available to its constituent communities subject to the access policies of each partner. The partners will assure that they have appropriate rights, ownership or permissions to share their data through the partnership. No transfer of ownership, copyright or otherwise, of the data or metadata shall endure to the partnership by acceptance of this MoU.

Any copyright in works the partnership obtains, creates or further develops will be owned by an individual member, and all members will be granted a perpetual, royalty-free license to use the work in any manner the member sees fit, including re-distribution for any non-commercial or commercial use by others.

8. Withdrawal and Removal of Membership

Partners may voluntarily withdraw membership by providing a 90 day written notice to the Steering Committee. The Data-PASS partnership may also remove members if commitments in this MoU are not met. The Steering Committee will make removal decisions, and will give 90 day written notice to the affected member prior to removal.

After withdrawal or removal, the partnership reserves the right, but is not obligated, to preserve existing metadata and data in the Data-PASS shared catalog and preservation network consistent with the objectives of the partnership and section 7 above. Retention of metadata and data in existence as of the time of member withdrawal or removal will be solely for the purpose of preservation; no transfer of ownership of the metadata or data shall occur.

9. Transfer Protocols

In the event that one of the partner institutions should no longer be able to retain the material they acquire, and no succession plan is already in place, the partner's materials will be transferred to one of the other partners to meet the objectives of the partnership. Both the original material and any additional material derived from the originals will be transferred.

Any partner that can no longer retain their material will provide the partnership with ninety (90) days written notice to provide for a reasonable transition. A determination regarding the most suitable partner to receive the material will be based on the subject matter and content of the materials as well as any economic encumbrances that may be involved in the transfer.

Ultimately, ICPSR would be the likely final repository for non-Federal materials and NARA the final repository for unique Federal materials. In the event that the university-based archives are no longer able to retain and preserve the content, they will agree to offer it to NARA. Materials acquired by the university-based archives that are not suitable for preservation by NARA because they are not related to Federal activity would be transferred to the Library of Congress in the most current format.

10. Costs

There are no regular membership fees for the Data-PASS partnership. Partners, however, are expected to contribute in-kind personnel time, as well as necessary computer equipment, to support the partnership objectives and to participate in the technology described in section 11 below.

11. Technology

The partnership agrees to develop and implement shared technology. Specifically, the partnership is committed to contribute to and maintain the following applicable to the objectives:

- Shared catalog (http://dvn.iq.harvard.edu/dvn/dv/datapass)
- Web site (http://www.data-pass.org)
- Intranet (https://wiki.hmdc.harvard.edu/DPP/)

As well, the partnership is committed to investigate and develop a system for replicated storage as relevant to the noted benefits of the partnership.

Data-PASS partners are supplied with the necessary source code, documentation, and procedures to participate in the technology infrastructure. The Steering Committee will approve technical, legal and procedural requirements for using each part of this infrastructure. Each partner is expected to comply with the requirements for use of the infrastructure, for purchasing and maintaining the equipment necessary to participate in the infrastructure, and to maintain the infrastructure.

12. Disputes

Any disputes amongst the Data-PASS partners will be handled by the Steering Committee.

13. Amendments

No amendment or changes to any provision of the MoU shall be valid unless in writing and signed by an authorized representative of each Data-PASS partner organization.

_____ _____

Name Date

Appendix B

Best Practices, Presentations, and Publications

Best Practices:

Articles of Collaboration
Original collaborative agreement for the Data-PASS partnership.

Content Selection
Information that provides general guidance to project staff as they seek to identify and review potential acquisitions.

Appraisal
Information that provides guidance on the appraisal of new materials identified by the partnership.

Acquisition
Standards for the work processes associated with the acquisition and common catalog tasks.

Metadata
Standards that are being used to allow for information so it can be included in a shared catalog and made available to shared systems.

Confidentiality
Policies to maintain data confidentiality of materials acquired for the Data-PASS project.

Data Security
Standards for the security of materials acquired for the Data-PASS project.

Guide to Handling Fragile Materials
Guidelines for Handling Fragile Materials (e.g. punched cards, magnetic tape, etc.) for Purposes of Data Recovery.

Data Deposit Agreement
Agreement that grants permission for Data-PASS, through its Partners, to archive and distribute data collections.

Appraisal and Acquisition of Actively Curated Collections
Guidelines for appraising and acquiring actively curated collections.

Data Identification Inventory and Checklist
Inventory and checklist for identifying and reviewing data collections.

Presentations:

2010

Crabtree, J. (2010, October). LOCKSS Auditing using the SAFE-Archive System. Presentation at the Educopia Institute "Private LOCKSS Networks: Community-based Approaches to Distributed Digital Preservation" Community Forum, Boston, MA.

Altman, M. (2010, October). Building PLN Communities. Presentation at the Educopia Institute "Private LOCKSS Networks: Community-based Approaches to Distributed Digital Preservation" Community Forum, Boston, MA.

Maynard, M. (2010, July). Data-PASS Partners: Plans for Moving Forward. Presentation at the 2010 Digital Preservation Partners Meeting, Washington, D.C. Retrieved August 30, 2010, from http://www.digitalpreservation.gov/news/events/ndiipp_meetings/ndiipp10/docs/July21/session01/maynard_breakout1_ndiipp2010_v2.ppt

Altman, M., Crabtree, J. & McGovern, N. (2010, July). Trustworthy Repositories, Organizations & Infrastructure. Presentation at the 2010 Digital Preservation Partners Meeting, Washington, D.C. Retrieved August 30, 2010, from http://www.digitalpreservation.gov/news/events/ndiipp_meetings/ndiipp10/docs/July21/session02/TRUSTWORTH_VOS_ALTMAN_CRABTREE_MCGOVERN.ppt

Altman, M. (2010, July). The Dataverse Network System and the SAFE-Archiving System. Poster presented at the 2010 Digital Preservation Partners Meeting, Washington, D.C.

Crabtree, J. (2010, June). Replicated & Distributed Storage Technologies: "Impact on Social Science Data Archive Policies. Presentation at the International Association for Social Science Information Services & Technology (IASSIST) 2010 Conference, Ithaca, NY. Retrieved August 18, 2010, from http://www.ciser.cornell.edu/IASSIST/program/pres/a3/IASSIST 2010 Distributed Storage Technologies.ppt

Adams, M.O. & Timms-Ferrara, L. (2010, June). USIA Office of Research Surveys, 1952-99, NARA -- Roper Center Collaboration: An Update . Presentation at the International Association for Social Science Information Services & Technology (IASSIST) 2010 Conference, Ithaca, NY. Retrieved August 18, 2010, from http://www.ciser.cornell.edu/IASSIST/program/pres/a3/IASSIST 2010 USIA presentation rev colors lois changes.ppt

Lyle, J. (2010, June). Building Partnerships Between Social Science Data Archives and Institutional Repositories. Presentation at the International Association for Social Science Information Services & Technology (IASSIST) 2010 Conference, Ithaca, NY. Retrieved August 18, 2010, from http://www.ciser.cornell.edu/IASSIST/program/pres/a3/ICPSR-IR_partnerships_20100602.ppt

Pienta, A. (2010, June). Retirement in the 1950s: Recovering The Cornell Study of Occupational Retirement. Presentation at the International Association for Social Science Information Services & Technology (IASSIST) 2010 Conference, Ithaca, NY. Retrieved August 18, 2010, from

http://www.ciser.cornell.edu/IASSIST/program/pres/a3/The Cornell Study of Occupational Retirement.ppt

Altman, M. (2010, June). How Collaborative Preservation Works. Presentation at the International Association for Social Science Information Services & Technology (IASSIST) 2010 Conference, Ithaca, NY. Retrieved August 18, 2010, from http://www.ciser.cornell.edu/IASSIST/program/pres/a3/altman_IASSIST2010.pdf

Altman, M. (2010, June). The Dataverse Network System and the Data-PASS Partnership. Poster presented at the International Association for Social Science Information Services & Technology (IASSIST) 2010 Conference, Ithaca, NY. Retrieved August 18, 2010, from http://ciser.cornell.edu/IASSIST/poster/ps11.shtm

Goodsell, L. & Teixeira, C. (2010, June). NARA-Roper Center Joint USIA Project. Poster presented at the International Association for Social Science Information Services & Technology (IASSIST) 2010 Conference, Ithaca, NY. Retrieved August 18, 2010, from http://ciser.cornell.edu/IASSIST/poster/ps13.shtm

2009

Altman, M. (2009, October). *Open Data: Legal and Institutional Models for Access and Preservation*. Invited Plenary Speaker at the International Conference on Preservation of Digital Objects (iPRES 2009), San Francisco, CA. Retrieved March 12, 2010, from http://maltman.hmdc.harvard.edu/presentations/ipres_v7.pptx

Gutmann, M. (2009, June). *Data-PASS Project Agreements*. Presentation at the June 2009 NDIIPP Partners Meeting, Washington, D.C. Retrieved January 18, 2010, from http://www.digitalpreservation.gov/news/events/ndiipp_meetings/ndiipp09/ docs/June25/Breakout-2-Org Agreements/datapass_agreements_20090625_new_template-v2.ppt

Crabtree, J. (2009, June). *Challenges and Opportunities in Repository Interoperation*. Presentation at the June 2009 NDIIPP Partners Meeting, Washington, D.C. Retrieved January 18, 2010, from http://www.digitalpreservation.gov/news/events/ndiipp_meetings/ndiipp09/ docs/June25/Breakout-3-Storage/2009_NDIIPP_breakouts_notes_session3.rtf

Granda, P., Thomas, D., Grasso, C., & Teixeira, C. (2009, April). *Common Workflows: Health and Social Science Data Curation Collaborations*. DigCCurr Conference 2009, Chapel Hill, NC. Retrieved June 18, 2009, from http://www.ils.unc.edu/digccurr2009/abstracts.html#granda

Crabtree, J. (2009, February). *A Community Approach to Preservation: "Experiences with Social Science Data"*. Community Approaches to Digital Preservation 2009 meeting, Atlanta, GA.

2008

Altman, M., Beecher, B., Maynard, M., & Crabtree, J. (2008, December). *A Platform for Auditable, Distributed, Asymmetric Archival Replication*. Presentation at the Coalition for Networked Information (CNI) 2008 Fall Task Force Meeting, Washington, D.C.

Pienta, A., Gutmann, M., Hoelter, L., Lyle, J., & Donakowski, D. (2008, August). *The LEADS Database at ICPSR: Identifying Important "At Risk" Social Science Data*. Roundtable paper presentation at the American Sociological Association Annual Meeting 2008, Boston, MA.

Timms-Ferrara, L. (2008, August). *Data-PASS*. Poster presentation at the American Sociological Association Annual Meeting 2008, Boston, MA.

Altman, M., & Beecher, B. (2008, July). *Multi-Archival Syndicated Storage Platform*. Presentation at the July 2008 NDIIPP Partners Meeting, Washington, D.C. Retrieved August 11, 2008, from http://www.digitalpreservation.gov/news/events/ndiipp_meetings/ndiipp08/ docs/plenary_custodians_altman.ppt

Pienta, A. (2008, July). *Restricted Data in the Social Sciences*. Presentation at the July 2008 NDIIPP Partners Meeting, Washington, D.C. Retrieved August 11, 2008, from http://www.digitalpreservation.gov/news/events/ndiipp_meetings/ndiipp08/ docs/session1_pienta.ppt

Gutmann, M. (2008, July). *The Data-PASS Partnership: Collaboration, Agreements, and More*. Presentation at the July 2008 NDIIPP Partners Meeting, Washington, D.C. Retrieved August 11, 2008, from http://www.digitalpreservation.gov/news/events/ndiipp_meetings/ndiipp08/ docs/session2_gutmann.ppt

Altman, M., & Crabtree, J. (2008, July). *Data-PASS Shared Catalog*. Presentation at the July 2008 NDIIPP Partners Meeting, Washington, D.C. Retrieved August 11, 2008, from http://www.digitalpreservation.gov/news/events/ndiipp_meetings/ndiipp08/ docs/session10_altman.ppt

Altman, M. (2008, July). *Living Collections, Ambient Data*. Presentation at the July 2008 NDIIPP Partners Meeting, Washington, D.C. Retrieved August 11, 2008, from http://www.digitalpreservation.gov/news/events/ndiipp_meetings/ndiipp08/ docs/plenary_content_altman.ppt

Lyle, J. (2008, July). *ICPSR Collection Development: Lessons Learned & Moving Forward*. Presentation at the July 2008 NDIIPP Partners Meeting, Washington, D.C. Retrieved August 11, 2008, from http://www.digitalpreservation.gov/news/events/ndiipp_meetings/ndiipp08/ docs/plenary_content_lyle.ppt

Maynard, M. (2008, May). *Solving Study Metadata Puzzles: Case Studies from Roper Center Reprocessing Activities*. Paper presentation at the International Association for Social Science Information Services & Technology (IASSIST) 2008 Conference, Palo Alto, CA. Retrieved July 30, 2008, from https://www.stanford.edu/group/ADS/cgi-bin/drupal/solving-study-metadata -puzzles

Crabtree, J. & Sheaves, D. (2008, May). *Impact of a Unified Catalog*. Poster presentation at the International Association for Social Science Information Services & Technology (IASSIST) 2008 Conference, Palo Alto, CA.

Altman, M. (2008, April). *Towards a Data Network for Integrated Social Science Research*. Presentation at the Digital Library Federation 2008 Spring Forum, Minneapolis, MN.

Altman, M. (2008, February). *Towards a Data Network for Integrated Social Science Research*. Paper presented at the Annual Meeting of the Association for the Advancement of Science (AAAS), Boston, MA. Retrieved March 21, 2008, from http://maltman.hmdc.harvard.edu/presentations/aaas_08.ppt

2007

Altman, M. (2007, November). *Social Science Data Preservation and Dissemination*. Poster presented at the Advancing NSDL Networks -- NSDL Annual Meeting, Arlington, VA. Retrieved March 26, 2008, from http://annualmeeting.nsdl.org/meeting/archives/2007/poster_abstracts.pdf

Gutmann, M.P., McGovern, N.Y., Beecher, B., & Raghunathan, T.E. (2007, October). *How Safe is Safe Enough when we Preserve Social Science Data?* Paper presented at the Third International Conference on e-Social Science, Ann Arbor, MI. Retrieved March 21, 2008, from http://ess.si.umich.edu/papers/paper150.pdf

Crabtree, J., Maynard, M., & Timms-Ferrara, L. (2007, October). *Developing Partnerships in the Social Sciences: Data Archives and Private Research Organizations*. Paper presented at the Third International Conference on e-Social Science, Ann Arbor, MI. Retrieved March 21, 2008, from http://ess.si.umich.edu/papers/paper162.pdf

Pienta, A. (Chair), Adams, M., Altman, M., Crabtree, J., Donakowski, D., & Maynard, M. (Panelists). (2007, August). *Data Preservation Alliance for the Social Sciences: A Model for Collaboration*. Session at the Society of American Archivists Annual Meeting, Chicago, IL. Retrieved March 21, 2008, from http://www.archivists.org/conference/chicago2007/chicago2007prog-Session.asp?event=2201

Altman, M. (2007, June). *Preserving Social Science Data Through Archival Collaboration*. Presentation at the Annual Meeting of the American Library Association, Washington, DC. Retrieved March 21, 2008, from http://maltman.hmdc.harvard.edu/presentations/ala_07.ppt

Altman, M. (2007, May). *Approaches to Data Dissemination and Preservation*. Presentation at the International Association of Social Science Information Technology Annual Meeting, Montreal, Canada. Retrieved March 21, 2008, from http://www.edrs.mcgill.ca/IASSIST2007/presentations/C1(2).ppt

Pienta, A. (Chair), Altman, M., Crabtree, J., Donakowski, D., & Maynard, M. (Panelists). (2007, May). *Data-PASS: Collaborating to Preserve At-Risk Data*. Session at the International Association of Social Science Information Technology Annual Meeting, Montreal, Canada. Retrieved March 21, 2008, from http://www.library.mcgill.ca/edrs/program.htm#F1

Crosas, M., & Treacy, R. (2007, May). *Harvards Dataverse Network: A JSF/EJB 3.0 Technology Data Sharing Solution on Java EE 5*. Presentation given at the JavaOne conference, San Francisco, CA. Retrieved March 21, 2008, from http://developers.sun.com/learning/javaoneonline/j1sessn.jsp?sessn=TS-4656&yr=2007&track=3

Altman, M., Crabtree, J., Donakowski, D., & Maynard, M. (Panelists). (2007, April). *Data Preservation Alliance for the Social Sciences: A Model for Collaboration*. Session at DigCCurr 2007: An International Symposium in Digital Curation, Chapel Hill, NC. Retrieved March 21, 2008, from http://ils.unc.edu/digccurr2007/papers/altmanEtal_paper_9-1.pdf

Altman, M., & Young, C. (2007, April). *Democratizing History On-Line: Moving from Microfilm to Megabytes*. Presentation at the American Association for History and Computing (AAHC) Annual Meeting, Providence, RI.

2006

Crabtree, J. & Donakowski, D.W. (2006, September). *Building Relationships: A Foundation for Digital Archives*. Presentation at the 1st European Workshop on the use of Digital Object Repository Systems in Digital Libraries (DORSDL), Alicante, Spain. Retrieved March 21, 2008, from http://www.lib.uoa.gr/dorsdl/dorsdl2006-crabtree.ppt

Altman, M. (2006, August). *Research and Data Support*. Poster presented at the American Sociological Association, Montreal, Canada.

Pienta, A.M. (2006, August). *Classic Social Science Studies to be Archived at ICPSR*. Poster presented at the American Sociological Association, Montreal, Canada.

Adams, M. & Granda, P. (Panelists). (2006, August). *New Archival Collaborations Between Traditional Government Archives and Academic Data Archives*. Session at the Society of American Archivists Annual Meeting, Washington, DC. Retrieved March 21, 2008, from http://www.archivists.org/conference/dc2006/dc2006prog-Session.asp?event=1783

Crabtree, J. & Donakowski, D.W. (2006, June). *Building Relationships: A Foundation for Digital Archives*. Digital Curation & Trusted Repositories: Seeking Success Workshop, in conjunction with the Joint Conference on Digital Libraries, Chapel Hill, NC. Retrieved March 21, 2008, from http://www.ils.unc.edu/tibbo/JCDL2006/Crabtree-JCDLWorkshop2006.pdf

Altman, M. & King, G. (2006, May). *A Proposed Standard for the Scholarly Citation of Quantitative Data*. Presentation at the International Association for Social Science Information Service and Technology Annual Meeting, Ann Arbor, MI. Retrieved March 21, 2008, from http://www.iassistdata.org/publications/iq/iq30/iqvol302king.pdf

Pienta, A. (2006, May). *The LEADS Database at ICPSR: Identifying Important Social Science Studies for Archiving*. Presentation at the International Association for Social Science Information Service and Technology Annual Meeting, Ann Arbor, MI. Retrieved March 21, 2008, from http://www.iassistdata.org/conferences/2006/presentations/F2_Pienta.ppt

Pienta, A.M., McNally, J., & Gutmann, M. (2006, March). The Life Cycle of Demographic Data at the LEADS Database at ICPSR. Poster presented at the Population Association of America, Los Angeles, CA. Retrieved March 21, 2008, from http://paa2006.princeton.edu/abstractViewer.aspx?submissionId=61527

Donakowski, D. (2006, January). *Data-PASS Content Identification and Partner Collaboration*. Poster presented at the January 2006 NDIIPP Partners Meeting, Washington, D.C.

2005

Arms, C., Altman, M., Carlson, M., Crabtree, J., Donakowski, D., & Maynard, M. (Panelists). (2005, August). *Data Archives: Building a Partnership to Preserve the Future of Social Research*. Workshop at the American Sociological Association Annual Meeting, Philadelphia, PA.

Donakowski, D. (2005, July). Data Preservation Alliance for the Social Sciences. Poster presented at the July 2005 NDIIPP Partners Meeting, Warrenton, VA.

Arms, C. (2005, May). *Data-PASS: A new effort to harvest our history (funder view)*. Presentation at the joint conference between the International Association of Social Science Information Technology Annual Meeting and the International Federation of Data Organizations, Edinburgh, Scotland. Retrieved March 21, 2008, from http://www.iassistdata.org/conferences/2005/presentations/A2Arms.ppt

Donakowski, D. (2005, May). *Data-PASS: A new effort to harvest our history (project update)*. Presentation at the joint conference between the International Association of Social Science Information Technology Annual Meeting and the International Federation of Data Organizations, Edinburgh, Scotland. Retrieved March 21, 2008, from http://www.iassistdata.org/conferences/2005/presentations/A2Donakowski.ppt

Timms-Ferrara, L. & Maynard, M. (2005, May). Partners in Preservation of Opinion Data: Developments of a Library of Congress Initiative. Poster presented at the American Association for Public Opinion Research (AAPOR) Annual Conference, Miami Beach, FL.

Donakowski, D. (2005, March). Archiving our Social Science Digital History. Presentation at the ECURE 2005: Preservation and Access for Digital College and University Resources Conference, Tempe, AZ.

2004

Adams, M.O. (Chair), Gutmann, M., Bollen, K., Carlson, M., Young, C., & Rockwell, R. (Panelists). (2004, May). *The Future of Social Science Data Archiving in the United States: A Discussion with Five Archive Directors*. Plenary at the International Association of Social Science Information Technology Annual Meeting, Madison, WI.

Publications:

Abrahamson, M., Bollen, K., Gutmann, M.P., King, G., & Pienta, A. (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.

Gutmann, M., Abrahamson, M., Adams, M., Altman, M., Arms, C., Bollen, K., Carlson, M., Crabtree, J., Donakowski, D., King, G., Lyle, J., Maynard, M., Pienta, A., Rockwell, R., Timms-Ferrara, L., & Young, C. (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(3).* Retrieved May 21, 2009, from http://muse.jhu.edu/journals/library_trends/v057/57.3.gutmann.html

Altman, M. (2009). Transformative Effects of NDIIPP, the Case of the Henry A. Murray Archive. *Library Trends, 57(3)*. Retrieved May 21, 2009, from http://muse.jhu.edu/journals/library_trends/v057/57.3.altman.html

Altman, M., Adams, M., Crabtree, J., Donakowski, D., Maynard, M., Pienta, A., & Young, C. (2009). Digital Preservation Through Archival Collaboration: The Data Preservation Alliance for the Social Sciences. *American Archivist, 72*(1). Forthcoming. Retrieved March 13, 2009, from http://maltman.hmdc.harvard.edu/papers/SharedPractices.pdf

Altman, M., Beecher, B., Crabtree, J., Andreev, L., Bachman, E., Buchbinder, A., Burling, S., King, P., & Maynard, M. (2009). A Prototype Platform for Policy-Based Archival Replication. *Against the Grain 21(2)*. Forthcoming.

Altman, M. (2008). A Fingerprint Method for Verification of Scientific Data. *Advances in Systems, Computing Sciences and Software Engineering*. Proceedings of the International Conference on Systems, Computing Sciences and Software Engineering 2007, 311-316. Retrieved March 13, 2009, from http://maltman.hmdc.harvard.edu/papers/unfcisse_rev_16_forthcoming.pdf

Altman, M., & King, G. (2007). A Proposed Standard for the Scholarly Citation of Quantitative Data. *D-Lib*, *13*(3/4). Retrieved March 21, 2008, from http://www.dlib.org/dlib/march07/altman/03altman.html

Altman, M., & King, G. (2007). An Overview of a Proposed Standard for the Scholarly Citation of Quantitative Data. *IASSIST Quarterly*, *30*(2), 18-19. Retrieved March 21, 2008, from http://www.iassistdata.org/publications/iq/iq30/iqvol302king.pdf

Green, A.G., & Gutmann, M.P. (2007). Building Partnerships Among Social Science Researchers, Institution-based Repositories and Domain Specific Data Archives. *OCLC Systems & Services*, *23*(1), 35-53. Retrieved March 21, 2008, from http://deepblue.lib.umich.edu/handle/2027.42/41214

King, G., (2007). An Introduction to the Dataverse Network as an Infrastructure for Data Sharing. *Sociological Methods & Research*, *36*(2), 173-199. Retrieved March 21, 2008, from http://gking.harvard.edu/files/dvn.pdf

Crabtree, J. and Donakowski, D. (2006). *Building Relationships: A Foundation for Digital Archives*. Paper presented at JCDL 2006 Workshop: Digital Curation & Trusted Repositories: Seeking Success. Retrieved

March 21, 2008, from http://sils.unc.edu/events/2006jcdl/digitalcuration/Crabtree-JCDLWorkshop2006.pdf

Gutmann, M., Schurer, K., Donakowski, D., & Beedham, H. (2004). The Selection, Appraisal, and Retention of Social Science Data. *Data Science Journal*, *3*, 209-221. Retrieved March 21, 2008, from http://www.jstage.jst.go.jp/article/dsj/3/0/209/_pdf