

# Data Preservation Alliance for the Social Sciences

**Content Selection.** The Data Preservation Alliance for the Social Sciences (Data-PASS) is in the process of systematically identifying and selecting digital social science data. We seek data that are:

- Theoretically and/or methodologically ground breaking
- Representative of a national, important regional, or hard to reach population
- Data collected as part of a policy evaluation
- Data cited as part of a seminal collection or tied to unrepeated or rare events

Once material has been selected for archiving, Data-PASS partners will begin the work of appraisal.

**Appraisal.** The partners will apply standard selection, or appraisal, criteria in reviewing data collections. In developing appraisal recommendations, the partners will answer the following questions:

- How significant are the data for research?
- How significant is the source and context of the data?
- Is the information unique?
- How usable are the data?
- What is the timeframe covered by the information?
- Are the data related to other data in the archives?

This framework will allow the partners to make consistent judgments that will help move the data collection into the acquisition and processing phase.

**Acquisition.** Effective data processing ensures that no matter what condition the materials were in when initially acquired, the material made available and preserved by the archive will be as complete, accurate, and well documented as possible. While some acquisition and processing decisions will be made on a case by case basis as needed, the primary guidelines that will be followed by each of the partners include:

*Acquisition of Study Data and Documentation:*

- Verify the contents of the data collection
- Perform initial confidentiality check
- Inventory and submit all materials for acquisition

*Processing Plan:*

- Determine the level of processing
- Resolve issues of content discrepancies
- Build the documentation set — study and file level metadata
- Package the study

When final quality control standards are verified, permission will be given for release and dissemination, with appropriate access controls in place. All materials created since the initial stages of acquisition will be secured for long-term preservation. The data collection's availability will be announced to the research community.

## Inter-university Consortium for Political and Social Research

As the lead institution for the Data-PASS project, ICPSR staff has guided the creation of our Articles of Collaboration, and appraisal guidelines and accompanying checklist, as well as drafting a version of the selection and acquisition guidelines that are to be used. A database is currently being created to track research leads. Leads originate from our contacts from conferences and organizations, and through monitoring additional sources (e.g., listservs for social science organizations). In addition to the presentations that have specifically focused on the Data-PASS project, an article coauthored by ICPSR staff and two presentations focusing on ICPSR have cited our role in the Data-PASS partnership, and have provided data leads and valuable information on current best practices.

## National Archives and Records Administration

NARA archivists have drafted the historical appraisal guidelines, resulting in the first project deliverable.

In recognition of the importance of preserving early social science data created by federal agencies, NARA made at-risk social science data one of the three focus areas upon which it would seek the transfer of historical electronic data "overdue" for transfer to the National Archives. NARA publicized this initiative in its bimonthly federal records management forums and semi-annual records management conference. The effort has so far resulted in the transfer of the following data:

- National Hospital Discharge Survey, 1979 to 2002 (Centers for Disease Control)
- National Ambulatory Medical Care Survey, 1975–1976 (Department of Health and Human Services)



## The Howard W. Odum Institute for Research in Social Science

The Odum Institute is focused on enhancing the collection of state polling data, identifying gaps in holdings from Harris Interactive, and pursuing data created by private research firms. We are working with our partner, the National Network of State Polls (NNSP), to develop Web-based tools, which will facilitate the transfer of datasets from state polling organizations to our archive. From Harris Interactive, we have received an initial set of data from nine surveys, which will provide a test-bed for developing and testing processes and procedures for evaluating data sets for inclusion in the Data-PASS collection. RTI International has recently solicited the submission of social science datasets from its subunits as a result of initial efforts to obtain data from private research firms.

## Harvard-MIT Data Center The Henry A. Murray Research Archive

As part of the Institute for Quantitative Social Science at Harvard University, both the HMDC and the MRA have contributed their unique perspectives to the creation of the partnership's guidelines and checklists. An extensive list of potential data has been identified, including data on minority and under-represented populations, and collections on American everyday family life. We released version 1.0.3 of the VDC system, adding wizards for creating studies and collections; a complete integrated help system; support for Fedora 3 Core; and numerous usability enhancements. We have also assisted the Odum Institute in setting up a VDC test site, and are working with them to map their current metadata to DDI. We have also conducted a presentation at the IASSIST in May describing joint work on federation with U.S. Census, citing our role in the Data-PASS project.

## Roper Center for Public Opinion Research

Staff at the Roper Center has been working with The National Opinion Research Center (NORC), reviewing their master list of studies. We have determined that there are approximately 250 NORC surveys that are not yet in our collection, and of those, 85 can be traced to their warehouse. NORC staff members are going through their paper records to see if documentation for the others can be found. We have also recently received a list of over 1,200 USIA studies from over 100 countries, and are going to begin the same process of matching it against our catalog to identify those that we do not have.

## Upcoming Presentation

### Data Archives: Building a Partnership to Preserve the Future of Social Research

The 100th Annual Meeting of the American Sociological Association (ASA)  
Philadelphia, Pennsylvania  
August 13–16, 2005

Sponsor:  
Research Workshops

Sunday, August 14 — 10:30am–12:10pm  
Building: Philadelphia Marriott

#### Abstract:

In 2004, six of the major social science data repositories in the United States joined together in a partnership with the Library of Congress to work on ensuring the long-term preservation of their holdings and of materials that they have not yet collected. In this workshop the leaders of that project will describe the services that they now offer, and discuss among themselves and with the audience the need to locate and preserve classic social science data, and the potential for future research based on them. The discussion will serve as an important starting point for ensuring that data used in the most important sociological discoveries of the past century are preserved for future research.

## Past Presentations

### Data-PASS: A Library of Congress Partnership

The 60th annual conference of the American Association for Public Opinion (AAPOR)  
May 12–15, 2005  
Miami Beach, Florida

### Data-PASS/NDIIP: A New Effort to Harvest Our History

Joint Conference between the International Association of Social Science Information Service & Technology (IASSIST) and the International Federation of Data Organisations (IFDO)  
May 24–27, 2005  
Edinburgh, Scotland