

INTRODUCTION

The Data Preservation Alliance for the Social Sciences (DataPASS) will achieve its goal of preserving electronic social science data through a process of identification and selection. The project's partners will attempt to identify the most significant digital social science data of the past seventy-five years, based on a variety of criteria. The content selection guidelines developed by the partnership and outlined in this document provide general guidance, and will be used in conjunction with the partnership's appraisal guidelines.

CONTENT SELECTION CRITERIA

DataPASS seeks to identify and select digital social science data that are classic or destined to be classic and at-risk of loss. Criteria to help define these data are as follows:

Criteria for Defining Classic Social Science Studies:

- Highly cited social science studies
- Studies conducted by highly cited social scientists
- Studies that are theoretically and/or methodologically ground breaking
- Studies based on a national sample, an important regional sample, or population historically underrepresented in research
- Data collected as part of a major policy evaluation
- Studies cited as a part of a seminal collection
- Studies tied to unrepeated or rare events

Criteria for Defining At Risk Social Science Studies

- Studies not currently in a permanent archive

SOURCES FOR IDENTIFYING DATA

What follows are examples of the data sources that we will vigorously investigate and from where we expect much of the data to come. We also identify the partnership organizations particular interests or expertise in each source.

1. Sources of University-Based Social Science Data

- Expert Staff Suggestions at ICPSR, the Murray Research Center, the Odum Institute, and the Roper Center will be solicited.

- Citation Databases including Social Sciences Citation Index & Institute for Scientific Information's database of citations will be reviewed by ICPSR.
- Internet Searches will be conducted by ICPSR to identify data being disseminated outside of archives. University-based data may be publicly available from data originators' web sites, but unlimited or continuous access is not necessarily guaranteed. The producers of these sites may have no long-term preservation plans; we will contact these producers and work with them to preserve their data.
- Survey Research Center, University of Michigan will be contacted by ICPSR to identify all major historic surveys conducted by this organization.
- The National Opinion Research Center (NORC) will be contacted by Roper to identify all major historic surveys conducted by this organization.
- The Roper Center's Traces Database will be reviewed by Roper staff. It contains 7,000 surveys referenced, but not found in the Roper archives.
- We will endeavor to form Ad Hoc Advisory Committees in major social science disciplines with leading researchers to assist us in the identification of important data collections in their respective fields. ICPSR is responsible for this activity.
- Suggestions by the academic research community will be sought by all DataPASS partners on an ongoing basis.

2. Sources of Federally Funded Research Data

- The CRISP (Computer Retrieval of Information on Scientific Projects) database will be reviewed by ICPSR for federally funded data awarded between 1972 and 2003 by the National Institutes of Health.
- The National Science Foundation (NSF) database will be reviewed by ICPSR for federally funded data awarded by NSF.

3. Sources of Federally Produced Research Data

- United States Information Agency (USIA) public opinion polls administered between 1953 and 1999 will be listed and reviewed by NARA and Roper to determine gaps in the holdings.
- Federal records, including those that are considered social science data, will be scheduled and archived according to Federal guidelines by NARA.

4. Sources of Political Process Data

- National Election and Polling Data will be identified by ICPSR.
- State and Regional Polling Data will be identified by Odum. The Odum Institute will cooperate with Dr. Ron Langley, Director of the National Network of State Polls, to locate and acquire promising state polls not already archived. Odum staff also will travel to conferences to enhance the Institute's ability to identify and archive relevant state poll data; Odum staff look to directly contact some of the major state-oriented polling organizations, as well.
- Local, State, and National Election Returns available on public web sites, but not already within ICPSR's archives, will be reviewed and captured by ICPSR.
- United States Congressional Roll Call Voting Records for the 105th, 106th, and 107th Congresses will be acquired by ICPSR to update this important series.

5. Sources of Private Organization Research Data

- RTI International will be contacted by the Odum Institute to identify all major historic surveys conducted by this organization.
- Harris Interactive will be contacted by the Odum Institute to identify and fill gaps in Odum's holdings of Harris bimonthly telephone poll data.
- The Odum Institute will endeavor to form an Advisory Committee of Private Research Organizations to help in developing criteria for identifying and selecting electronic data from private organizations.

6. Sources of Vulnerable Data in Specialty Archives

- Specialty Data Archives will be listed by ICPSR. The sources for this list include: a review of social science departments in universities (specialty archives are often attached to university departments), contacting professional organizations, and Internet Searches.

RESPONSIBILITIES

The identification and selection process implemented by Data-PASS will generally be decentralized with each archive pursuing data that best represent its content area of specialization. This decentralization allows each partner to leverage their distinct

capabilities in specific kinds and sources of data. Each partner will then submit information on identified sources to a master database of all potential data collections.

As data collections are identified, selection of available data will be centralized through the work of the Operations Committee. The data's selection will be based on appraisal guidelines including, significance of the data to the research community, significance of the source and context of data, uniqueness and usability of the data and so on. (For more on this process, please review our Appraisal Guidelines.) The committee will review the available information in the database about each individual data collection and *codify* the information via a checklist. This checklist will become part of our records, and will be accessible for review. DataPASS partners will attempt to archive the highest rated studies.

Once data are identified for acquisition, the process will again become more decentralized. Following fundamental acquisition and processing guidelines, as agreed upon by the Operations Committee, the responsibility of acquisition, processing, and preservation will be assumed by the partner organization best suited to complete the tasks. The partners will be allowed and encouraged to complete additional tasks, as warranted.

The Inter-university Consortium for Political and Social Research (ICPSR) will provide centralized leadership and oversight of the activities of the university-based partners and will coordinate with the National Archives and Records Administration (NARA) so that efforts are not duplicated across the archives. ICPSR is also leading the effort to review the NSF database of funded data projects. While the initial review has produced promising results, it has also drawn attention to challenges that the partnership will face in our selection activities. For the most part, historic records are incomplete. The lack of complete information in the NSF database, as well as other sources, will undoubtedly prompt additional discussions among both the Steering and Operations Committees. Attention will be required to the balance between using resources to obtain additional information on incomplete records and using these resources to acquire and preserve collections with more complete initial information available.

SUMMARY

The sources provided in these guidelines are our starting point. Just as the expertise and experiences from among the partners' staff allow us to self-identify important data collections, data will also be identified through communication with other experts and eminent researchers across the social sciences.

Regardless of how the data is identified, we recognize that some of these studies, upon closer investigation, may not be available to us. Some data may not be accessible or may not be central to the focus of this project. However, every effort will be made to acquire as many of these studies as possible.

While these guidelines provide a foundation on which to build our collections of digital social science data, we expect that some guidelines will change over time according to our experiences and increasing expertise. Such a dynamic process is both expected and encouraged.

Approved by the Steering Committee on December 12, 2005