

Data-PASS : Its Origins and Accomplishments

**A Presentation for the Georgetown Law Library's
Symposium on Big Data and Big Challenges for Law
and Legal Information, January 30, 2013**

Margaret O. Adams, Supervisory Archivist, Retired
Research Services, Electronic Records Section
National Archives & Records Administration

Data-PASS: initially a NDIIPP Project

Goals

- **Ensure valuable and potentially at-risk social science data are**
 - Identified
 - Acquired
 - Preserved
- **Establish a basis for collaboration into the future**

The DataPASS Partners: Major American Social Science Data Archives

- **Archival Electronic Records Operations, Research Services – Washington, D.C., National Archives & Records Administration**
- **Harvard-MIT Data Center**
- **ICPSR -- Inter-University Consortium for Political and Social Research, based at the University of Michigan**
- **Murray Research Archive, Institute for Quantitative Social Science, Harvard University**
- **Odum Institute for Research in Social Science, University of North Carolina, Chapel Hill**
- **Roper Center for Public Opinion, based at the University of Connecticut**
- **UCLA Social Science Data Archive**

Shared Goal of the Partners

- Long-term accessibility to datasets for social science research
 - Accessibility depends upon identifying, appraising, acquiring, describing, and preserving digital data
 - These are functions that are basic and shared by academic data archives and traditional archival institutions

Partnership Goals

- Déjà vu: the 1960s – 1970s
 - Internationally: ventures in collaboration sponsored by UNESCO through the International Council of Social Science Data Archives
 - Nationally [U.S.]: work of the NSF-sponsored Council of Social Science Data Archives (CSSDA), 1965 - 1970
 - Internationally: 1974 founding of IASSIST (International Association for Social Science Information Service & Technology, www.iassistdata.org)

Forty Years Later

- In the U.S. DataPASS is beginning to reach 1960s collaboration goals of the CSSDA
 - Aided by 40+ years of evolution and experience in both social science and technology
 - Based on a recognition of a national, shared interest in long-term accessibility to primary digital data from social science research and federally created digital data

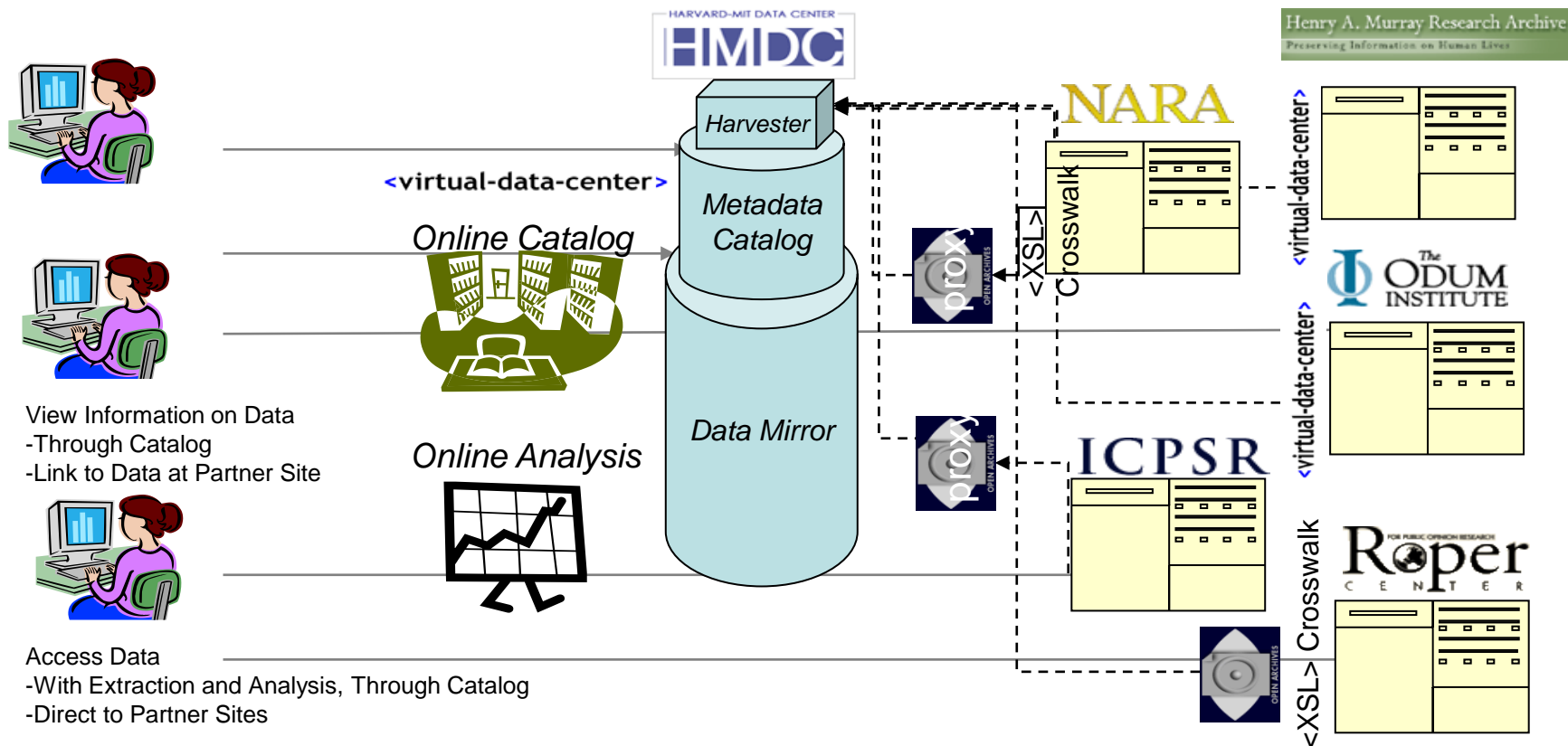
Contributions of DataPASS: Best Practices

- DataPASS Appraisal Policy
 - DataPASS colleagues collaborated in drafting guidelines
 - Approved by the Operations Committee, 2/9/2005
 - Reflects traditional archival appraisal criteria
 - Focuses on informational, and to a lesser extent, evidential values in appraising social science data
 - See, <http://data-pass.org/best-practices.jsp>

The DataPASS Shared Catalog: its reality actualizes a goal first formulated by the CSSDA

- **Serves as a symbolic link of past and present**
- **Based on national descriptive standards for digital data**
- **A common catalog for data holdings from major U.S. social science academic data archives, plus the descriptions of data records of the U.S. National Archives**
- **<http://data-pass.org/call.jsp>**

Architectural Overview



Contributions of DataPASS: Preservation Policies

Academic partners built a prototype for a storage platform for policy-driven distributed replication of digital holdings

Developed open source tools that can be used generally for content replication services

See: <http://data-pass.org/syndicated-storage.jsp>

Contributions of DataPASS - Content

National Archives and Roper Center collaboration built mutually comprehensive collections of USIA worldwide survey data files, 1950s to 1999 – close to 2000 data files now preserved at NARA; a subset at Roper Center

Pre-DataPASS, NARA's and the Roper Center's holdings: both unique and overlapping files: substantial collaboration of staffs involved – Roper undertook reformatting to assure long-term usability

Contributions of DataPASS: Citation Standards

- The Data-PASS partners are committed to promoting standards and improving practices for the citation of data.
- Developed guidelines and best practices for citing social science research data in order to promote vigorous and consistent attribution of datasets.
- See, <http://data-pass.org/citations.html>

DataPASS Intangibles

Value of Collaboration and Partnership as part of a larger national digital preservation effort

Opportunity for Demonstrating Commonalities between Traditional and Data Archival Work

Realities of Contemporary Virtual Communities
Actualize the Ideas of the Past

Potential Relevance for the Archival and Library Professions

Digital Data are created through the research activities of faculty and students throughout the country: what is the long-term future for these materials?

Many campuses have archives and/or manuscript repositories; some have local data archives or social science data centers

... Collaborate

... and Consider the value of a national shared virtual repository for digital social science data

Bibliography:

Altman, Micah, *et.al*, “Digital Preservation through Archival Collaboration: The Data Preservation Alliance for the Social Sciences,” *The American Archivist* 72 (Spring/Summer 2009), 170-184.

Gutmann, Myron P., *et. al*, “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, pp. 315-337.

Thank you.

<http://data-pass.org/>