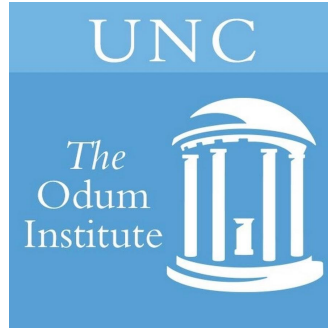# Library Data Archive

Providing archive and data services since 1961
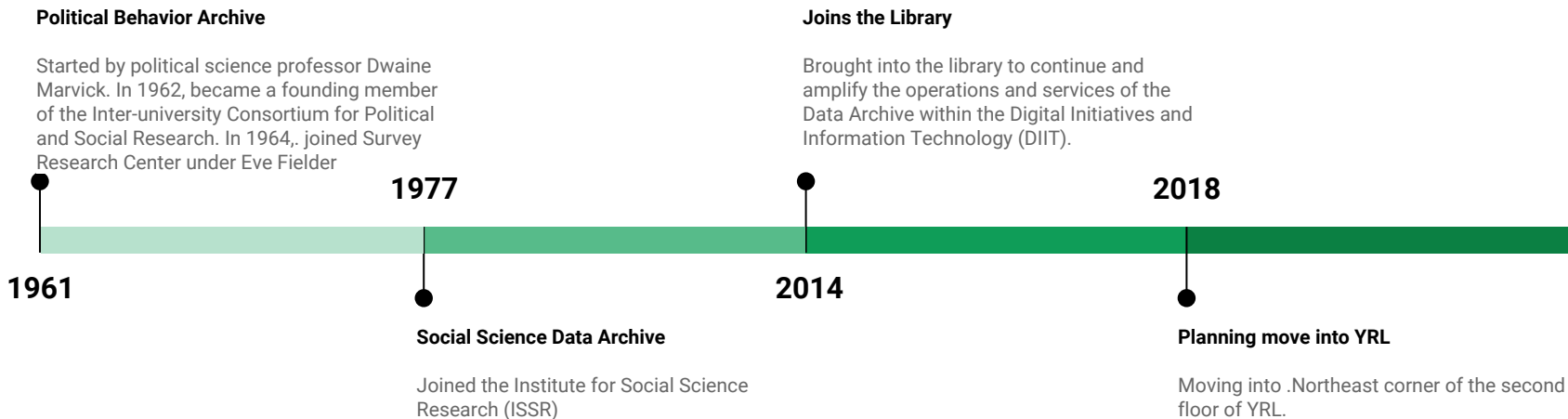
November 20, 2017

What is a Data Archive?

What do data archives typically do?

- Data collection, management and preservation
- Research support through the data lifecycle
- Data and computing support
- Instruction on, tools, methods, data management and analysis
- Reflects empirical nature of social science research and the advent of centralized computing and statistical computing

SSDA

Social Science Data Archive

Political Behavior Archive

Started by political science professor Dwaine Marvick. In 1962, became a founding member of the Inter-university Consortium for Political and Social Research. In 1964,. joined Survey Research Center under Eve Fielder

Joins the Library

Brought into the library to continue and amplify the operations and services of the Data Archive within the Digital Initiatives and Information Technology (DIIT).

1977

2018

1961

2014

Social Science Data Archive

Joined the Institute for Social Science Research (ISSR)

Planning move into YRL

Moving into .Northeast corner of the second floor of YRL.

Table 1.1 *The earliest social science data libraries, centres and archives*

Source: http://goo.gl/gQJrK8, reprinted with permission of Luis Martinez.
Several social science data archives, centres and libraries were in place by the 1970s, and many more were established by the end of the 20th century.

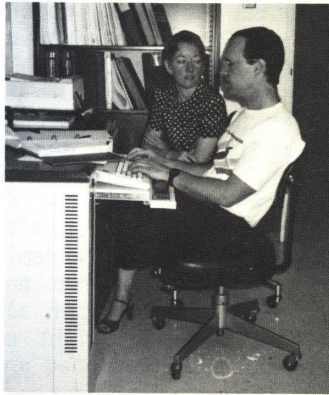| Institution	Year	Type	Country	Other names
Roper Center	1957	Data centre	USA	
Berkeley	1958	Data library/ service	USA	
GESIS – Leibniz Institute for the Social Sciences	1960	Data centre	Germany	Central Archive for Empirical Social Research
UCLA Data Services	1961	Data library/ service	USA	
ICPSR	1962	Data centre	USA	
Data Archiving and Networked Services (DANS)	1964	Data centre	Netherlands	Steinmetz Archive
Social Science Data Archives, Department of Sociology and Anthropology, Carleton University	1965	Data library/ service	Canada	
Princeton Data Service	1965	Data library/ service	USA	
Data & Program Library Service, University of Wisconsin-Madison	1966	Data library/ service	USA	Data & Information Services Center (DISC)
The Institute for Policy and Social Research	1966	Data centre	USA	
UK Data Archive	1967	Data centre	UK	UK Data Service
Data Library/Service RAND Corporation	1968	Data library/ service	USA	
Electronic Records Division, US National Archives	1968	Data centre	USA	NARA Data Archives Program
Odum Archive, University of North Carolina	1969	Data centre	USA	The Odum Institute

Role of the Data Archive in Soc. Sci. Div.

- Supported social science research through all stages of a project, including:
  - hypothesis or area of study development
  - grant and funding activities
  - data collection, analysis, management, and communication
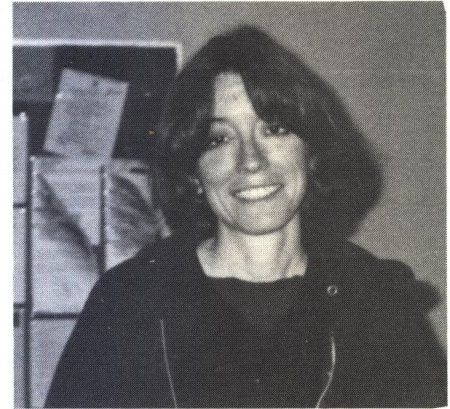 - long term preservation of research data/results/artifacts

The Social Science Data Archive: Access to Machine-Readable Data

"It must be here somewhere." Elizabeth Stephenson, Data Archivist, looks through the Archive's data tape collection.
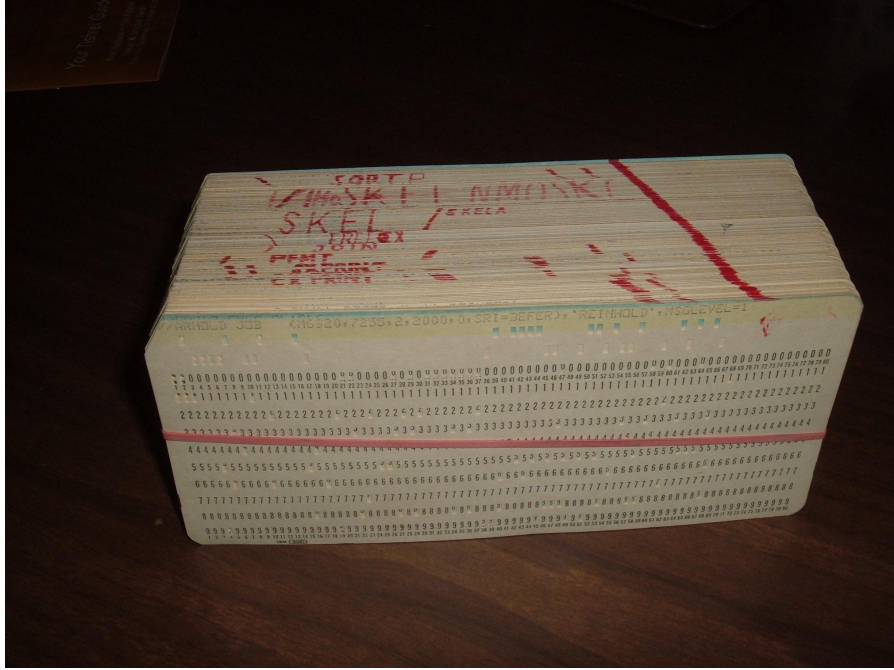

Martin Pawlocki, Programmer, maintains and develops the administrative software for the Institute and provides assistance to staff and Archive users.

Gina Jamison, Technical Assistant in the Archive, will be receiving her Master of Fine Arts and leaving ISSR in June.

DATA ARCHIVES LIBRARY

The Data Archives Library has re-located from Franz Hall to ISSR's new facility in the Gayley Center Building

ISSR welcomes a new Programmer/ Technical Consultant, Shirley Lane. Ms. Lane has been the consultant for the machine-readable version of the 1970 Census at UCLA's Research Library. She joins the staff as part of a cooperative effort between the Research Library and the Data Library to provide informat and reference services utilizing ma___dable data files. When the Dat___ begins receiving the 198___ ___ne's talents will b___

On-Line Data Files

The Archive presently maintains a selection of heavily used files on the SSC 4381 (Datalib). Because of the limited space available on the 4381 for such storage, and the need to rotate what is stored, most of these files will be moved to the OAC 3090 and the Archive will release most of this space to users of the 4381. Because of their frequent use in classes, modules for accessing CITIBASE and data from the IMF will be available on both the 3090 and the 4381.

Technical Consulting

The Archive will assist users to mount tapes or cartridges, access on-line files, or solve simple technical problems on whatever machines or operating systems preferred. Information on how to access tapes or online files will be provided in

DATA ARCHIVE SEMINAR
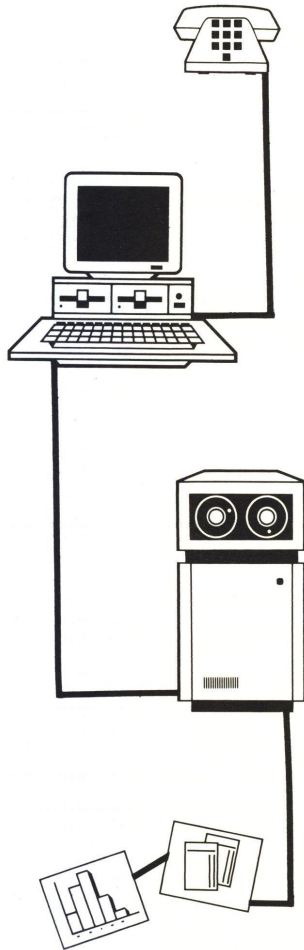
by Charlie Recksieck

The Social Science Data Archive (SSDA) will be holding a user seminar on the use of information files in gender research. The seminar, entitled "The Use of Data Files for the Study of Women," is scheduled for February 9th from 10:00 to 11:30 a.m. in 9383 Bunche.

Eli___eth St___enson ___chivist and

SPIN

Often, just finding infor-___tion concerning grant oppor-___nities can be the biggest ___arrier to conducting funded research. This year the ISSR Data Archive and the Center for the Study of Evolution and Origin of Life (CSEOL) will begin to assist researchers in discovering sources of funding for research, fellowships, travel, training and atten-

Collections

- Approx. [600 datasets](#) (some unpublished, ask us)
- Mostly from surveys, public opinion polls, and de-identified administrative records.
- Include studies on:
  - Political, social & economic attitudes and behavior
  - Studies of health status and health outcomes
  - International surveys of occupational and social mobility
  - Los Angeles County Social Surveys & California Polls
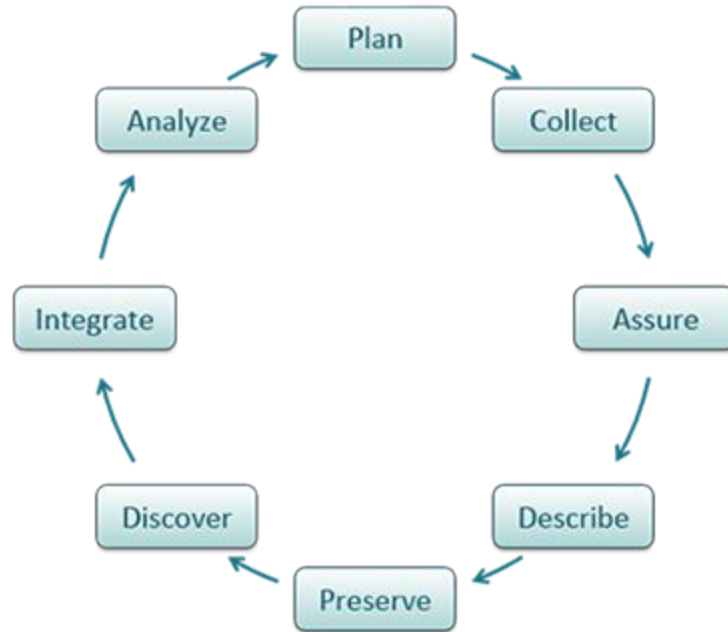
SSDA in the Library

Data is the central currency of science, but the nature of scientific data has changed dramatically with the rapid pace of technology. – Ted Hart (@emhart), et. al., Ten Simple Rules for Digital Data Storage.doi:10.1371/journal.pcbi.1005097

Access to the computational steps taken to process data and generate findings is as important as access to data themselves. – Victoria Stodden, Enhancing reproducibility for computational methods. Science 09 Dec 2016 : 1240-1241. 10.1126/science.aah6168

The Data Archive will (continue to) provide research data services in three interrelated areas:
1. Consulting and data support;
2. Archive and repository services; &
3. Educational program.

The Data Archive will continue to support all phases of the data lifecycle:

Also support research innovation

- Machine learning
- Network analysis
- Text analysis
- Cloud computing and storage
- Web data and APIs
- Data visualization
- Programming service and code review

Archive and Repository Services

Continue archival services

- Data curation support
- Data management best practices support and training
- Data stewardship

Add data publishing & sharing

- Data Den and Dataverse repositories
- Curating for reproducibility
- Working with data at scale
- Be responsive to interoperability needs (ORCIDs, APIs, Preregistration)

In a 2016 survey of 704 NSF Biological Sciences Directorate principle investigators… **nearly 90% indicated they are currently or will soon be analyzing large data sets**. …. the **most pressing unmet needs are training in data integration, data management**, and scaling analyses for HPC … This portends a **growing data knowledge gap** in biology and challenges institutions and funding agencies to **redouble their support for computational training** in biology.

New knowledge and skills will be needed to make effective use of new system architectures and software. "Hybrid" disciplines such as **computational science** and **data science** and interdisciplinary teams may come to **play an increasingly important role**. Keeping abreast of a **rapidly evolving suite of relevant technologies** is challenging for many computer science programs ...

Other topics for instruction

- Concepts of reproducible research
  - Application of concepts in a tool (R, Stata, Python or Open Science)
- Data publishing
- Data cleaning topics in OpenRefine and R
- Text mining in R
- Data visualization
- Research project management, package/dependency management and capturing compute environments
- HPC and cloud computing

Fall Quarter Offerings

2017

R Workshop - Master of Social Science	2017/11/28
Python Workshop - Master of Social Science	2017/11/16
SQL Workshop - Master of Social Science	2017/11/02
Hands-on Introduction to Literate Programming	2017/10/27
Open Science Framework for Research Workflows	2017/10/18
Data Archive Class Presentation for Masters of Social Science	2017/10/17
Software Carpentry - UCLA	2017/10/16
Data Archive Class Presentation Psych M144/Stats M154	2017/10/05
Library Carpentry Workshop	2017/09/30
Library Carpentry South Africa	2017/09/30
Open Refine Part 2 and BYOD	2017/09/22

Thank You

ucla-data-archive@googlegroups.com