

Mainframe to Dataverse

1961 to 2019

UCLA Social Science Data Archive to dataverse.ucla.edu

1961: UCLA Political Behavior Archive

The **Political Behavior Archive** was started by Political Sci Professor **Dwaine Marvick**.

1962: Inter-university Consortium for Political and Social Reserach (ICPSR)

UCLA becomes a founding member of **ICPSR**.

1964: Survey Research Center

The Archive joined the **Survey Research Center**. Under Center Director, **Eve Fielder** the Archive would curate and archive unique material created by the Survey Research Center and material requested or purchased by researchers



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

1977: Social Science Data Archive

- The Archive (now **Social Science Data Archive**) and **Survey Research Center** become part of the **Institute for Social Science Research (ISSR)**.
- **Elizabeth (Libbie) Stephenson** hired as Director and Archivist.
- The collection eventually would grow to 667 datasets and ~5000 files.

2014: Acquired by UCLA Library

UCLA Library acquisition of the Data Archive to support research and data services for the campus.

Moving from the Social Sciences Division to the Library required a change in storage location. Libbie Stephenson chose Dataverse, hosted at Harvard Dataverse.

2016 Libbie Stephenson retires.

2018 **Tim Dennis** hired as Director in 2018.

2019: UCLA Library Data Science Center

The **Social Science Data Archive** is re-branded as the **UCLA Library Data Science Center**.

Mainframe Era

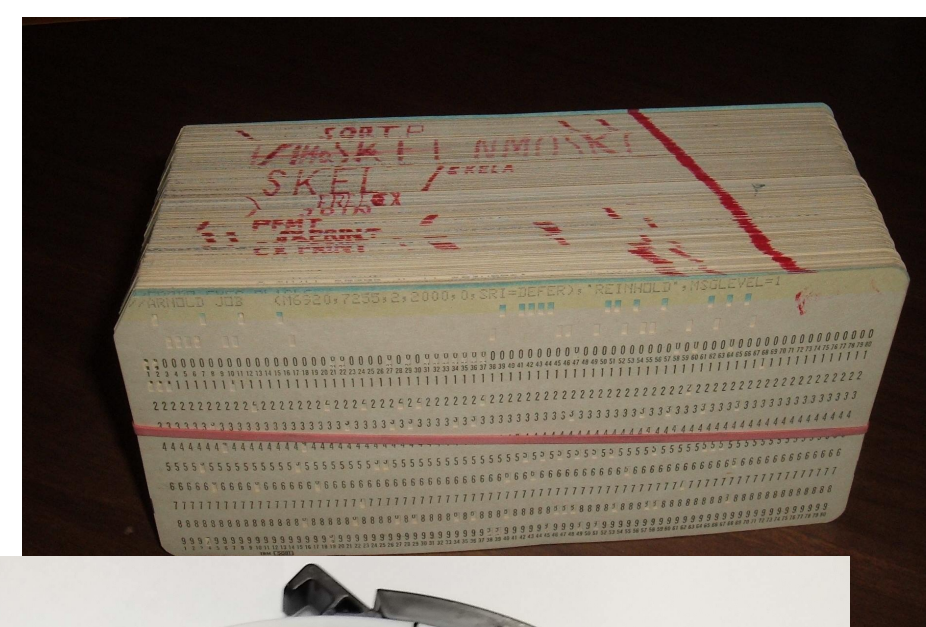
- Data stored on magnetic tape. Subjects included US Census data, vital Statistics, natality data, mortality data, researcher requested datasets and unique studies created at UCLA
- Analysis required the data to be loaded onto the mainframe via *Job Control Language (JCL)*. The datasets were stored at the archive when not being loaded on the mainframe.
- Access and discovery usually required a reference interview.



UCLA Mainframe:
NOT the original mainframe. A later model, IBM 3090 c. 1987-88.

Format Migration

- In the mainframe era data was stored externally and loaded as-needed.
- As storage devices improved or evolved this would require multiple format migrations to insure material remained accessible. (ie. Obsolete media may not be readable.)
- As well as storage formats migration might also mean:
 - Character encoding: EBCDIC to ASCII
 - Different systems: IBM Mainframe to Unix
- As personal computers evolved more storage formats were added to the mix.



Hosting with Local Storage (Post-Mainframe)

- 1996 UCLA Mainframe hosting ends.
- 1997 first online catalog for the Data Archive collections.
 - Developed in-house
 - Data Archive staff responsible for coding and maintenance
 - 1st version **dBaseIII**, static web pages
 - 2nd version **PHP/MySQL**
- Social Science Computing provided host machine and scheduled backups.
- Improved access to collection
- Initially datasets were still considered too large for personal computers in the 1990s. Researchers were required to have accounts on the Social Science Computing lab Unix machine.
- As personal computers evolved to manage larger datasets patrons could search the catalog, authenticate via Shibboleth, download datasets and documentation.



Complete Catalog

This is the catalog of all studies at the Data Archives. You may Search by broad Index terms that we use to categorize the collection, by words in study Titles; by the most recently received studies or Updates; or you can do a keyword Search. A keyword Search will look for words that appear anywhere in the information about a study. You should also search the ICPSR catalog, where you can narrow your search and find studies more precisely. For questions, please contact the Archive.

Index Titles Updates
Search

Data Archives Home
Copyright 1996 Institute for Social Science Research. Last Updated: 12-12-2003 - Stephenson.

Moving to Dataverse

2014: Harvard Dataverse hosted Archive

Dataverse chosen by Libbie Stephenson to replace the local repository and in-house catalog.

Her reasons for choosing Dataverse included:

- Archive staff no longer solely responsible for all code and maintenance
- Dataverse community and support
- Connection to Social Sciences and Social Science journals
- Support for DDL Metadata schema
- Ability to manage restricted data
- Scalable

2019: Locally Hosted Dataverse

In order to meet expanding archiving and repository needs on campus the Data Science Center planned a local instance of Dataverse.

Choice of platform: Amazon Web Services (AWS)

- EC2, t2.large
- RHEL
- S3 buckets for data and cross-region replication

- Install using IQSS/dataverse-ansible script
 - **test.dataverse.ucla.edu**
 - **dataverse.ucla.edu**
- Authentication via UCLA Shibboleth
 - Production system accounts are limited to UCLA faculty, researchers, staff and students
 - Non-UCLA accounts can be added as needed.

- Social Science Data Archive metadata harvested from Harvard Dataverse



<https://dataverse.ucla.edu/>
datascience@ucla.edu

Template from Colin Purrington: <https://colinpurrington.com/tips/poster-design>