

UCLA Library Data Archive

Doing Data since 1961

Fall Quarter 2017

1 / 35

Today's Topics

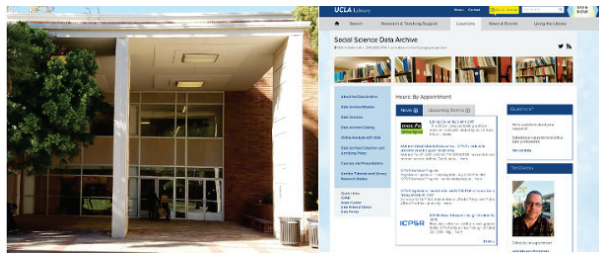
- UCLA Library Data Archives: who we are and what we do
- How to search for and access data for research
- Resources, tools and support available to students at UCLA
- Publishing and Data Management Plans

2 / 35

UCLA Library Data Archive

(formerly: Social Science Data Archive)

1120 Rolfe Hall



<https://www.library.ucla.edu/location/social-science-data-archive>

3 / 35

Data Archives: Our Services and the Data Life Cycle

The Data Archive provides data services that are tied to the data life-cycle.

"Data archiving is a process, not an end state where data is simply turned over to a repository at the conclusion of a study. Rather, data archiving should begin early in a project and incorporate a schedule for depositing products over the course of a project's life cycle and for the creation and preservation of accurate metadata, ensuring the usability of the research data itself. Such practices would incorporate archiving as part of the research method." Preserving Research Data, Jacobs and Humphrey (2004)

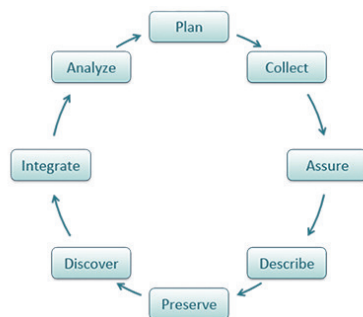


illustration: DatOne Data Life Cycle

4 / 35

Data Archives: Our Services and the Data Life Cycle

<https://www.library.ucla.edu/location/social-science-data-archive/data-services>

Plan

- Review data sources
- Help investigate archiving issues, costs, consent and disclosure risks
- create a data management plan

Collect/Assure

- Project management and file organization
- Quality Assurance for collected data and acquisition workflows
- Consultation on restricted data access and security

Describe

- Assist with filenames, standard terminology and data dictionaries
- Document analysis and file manipulations
- Help identify appropriate standards

5 / 35

Data Archives: Our Services and the Data Life Cycle

Preserve

- Help with backing up your Data
- Help with deciding what data to preserve
- Choosing stable file formats
- Help identifying suitable repositories

Integration

- Consider compatability
- Document steps
- Capture the provenance of sources

Analysis

- Help identify appropriate software
- Consultation on coding best practices

6 / 35

Data Archives: Our Services and Software

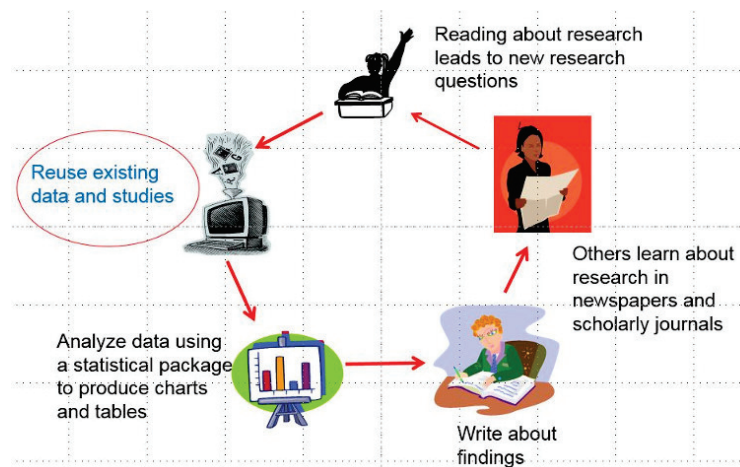
- Provide training: [Software/Library carpentry](#) / [Software Carpentry R-language](#)



- Host events and lectures: [Data Archive News and Upcoming Events](#)
- Statistical Consulting: Refer to [UCLA Statistics Consulting Center](#)

7 / 35

Role of Data in Research: Data Reuse



8 / 35

Finding Out About Data Sources

- UCLA Library Data Archive
- Colleagues
- Public media (newspapers, news)
- Social media (twitter, etc)
- Scholarly literature
- Government reports and documents
- Archives and project websites
- Codebooks, questionnaires
- Exploration of datasets

9 / 35

Data Sources

- ICPSR: <http://www.icpsr.umich.edu/icpsrweb/ICPSR/>
- UCLA Library Data Archive Catalog: https://dataverse.harvard.edu/dataverse/ssda_ucla
- UCLA Library DataDen: <https://dataden.library.ucla.edu/jspui/>
- Government Sites, Census Data: <https://www.census.gov/>
- IPUMS (Integrated Public Use Microdata Series): <https://www.ipums.org/>
- Public Policy Institute of CA: <http://www.ppic.org/>
- NBER (National Bureau of Economic Research): <http://www.nber.org/data/>
- CDC/NCHS (Centers for Disease Control and Prevention/National Center for Health Statistics): <https://www.cdc.gov/nchs/surveys.htm>
- Pew Research Center: <http://www.pewresearch.org/>
- Data.gov: <https://www.data.gov/>
- World Health Organization: <http://www.who.int/healthinfo/statistics/en/>
- International Monetary Fund: <http://www.imf.org/en/Data>

10 / 35

Search Strategies and Defining Your Research

- General to specific; broad to narrow
- Names of investigators; names of studies
- Studies described in literature, news and websites

Geography	Physical areas	Government units	Political boundaries	Other – school districts, etc.	
Time period	Current		Historical		Over time
Population & Units of analysis	Gender, age, etc.	Race, ethnicity, ancestry, etc.	Persons	Households	Other – schools, hospitals, etc.
Context	What is your hypothesis? What will the data tell you? How will you interpret it?				
Variables in the dataset	What variables do you need to answer your research question?				

How do I know this study useful?

- Raw data are **not** eye-readable
- To decipher data you will use **Questionnaires and Codebooks**

Dump of Data -- This is an example of a raw data file.

[illegible]

What is a Questionnaire?

- A **questionnaire** is a *research instrument* consisting of a series of *questions* and other prompts for the purpose of gathering information from respondents.
- Questionnaires cover separate topics such as:
 - Preferences (e.g. political party)
 - Behaviors (e.g. food consumption)
 - Facts (e.g. gender)
- Questionnaires can include indexes or scales:
 - Latent traits (e.g. personality traits)
 - Attitudes (e.g. towards immigration)
 - An index (e.g. Social Economic Status)

13 / 35

Questionnaire: ABC New Poll

TNS Job #175374
Horsham, PA July 23, 2008
FINAL

← **Date of poll**

ABC News/Discovery Channel/Stanford University Poll #1067: Global Warming Poll

Title of Poll → Hello, I'm (NAME), calling for the ABC News public opinion poll. We're not selling anything, just doing an opinion poll on interesting subjects for the news. Out of all the people age 18 or older who are AT HOME RIGHT NOW, can I please speak to the YOUNGEST (male/female)? (IF NO MALE/FEMALE IN THE HOUSEHOLD, ASK: Then may I please speak to the YOUNGEST (female/male)?)

Question text → 1. How would you rate the condition of the natural environment in the world today --
excellent, good, fair, poor, or very poor?

1 Excellent
2 Good
3 Fair
4 Poor

5 Very poor
DK/No opinion
NA/Refused

← **Response categories and values**

14 / 35

What is a Codebook?

- An overview of the study or dataset
- Variable names and descriptions
- Column locations (mostly this is for raw data)

May also include:

- Test of questions if the data comes from a survey
- Example of the Questionnaire
- Detail on who responded to the survey

15 / 35

Codebook: General Social Survey

Source: General Social Survey, 1972-2004

Variable name	HLTHINFO		R SOUGHT HEALTH INFORMATION PAST YEAR		
Question text	Text of this Question or Item 792. In the past year, that is, between (INSERT CURRENT MONTH) 1999/2001 and (INSERT CURRENT MONTH) 2000/02 have you looked for information about a health concern or medical problem?				
	% Valid	% All	N	Value	Label
	49.8	3.5	1,628	0	NOT AT ALL
	20.6	1.5	675	1	1 OR 2 TIMES
	17.1	1.2	559	2	3-5 TIMES
	12.5	0.9	408	3	6 OR MORE TIMES
		0.0	13	8	DK
		0.1	27	9	NA
	92.9	43.200	.		(No Data)
	100.0	100.0	46,510		Total
Properties					
Data type:	numeric				
Minimum code defined as valid:	0				
Maximum code defined as valid:	3				
Missing-data codes:	-1,8,9				
Record/column:	1/3360				

Number of responses

Values and labels

16 / 35

Q	VAR	INDEX	SQL
1	119	05	7-8
<p>WOW, I'D LIKE TO ASK THE SAME QUESTION ABOUT AHEAD YOU LIVE. HOW LONG HAVE YOU LIVED IN THE LOW RUCKER AREA?</p> <p>00. less than 1 month 01. 1 month to less than 2 etc. 11. 11 months to less than 12 12. less than 1 year--subset of months not specified 13. 1 year to less than 2 etc. 96. 96 or more years 97. refused 98. DK 99. NA</p>			
Q	VAR	INDEX	SQL
3	120	05	9-10
<p>HOW LONG HAVE YOU LIVED AT YOUR PRESENT HOMESET?</p> <p>00. less than 1 month 01. 1 month to less than 2 etc. 11. 11 months to less than 12 12. less than 1 year--subset of months not specified 13. 1 year to less than 2 etc. 96. 96 or more years 97. refused 98. DK 99. NA</p>			

Dump of Data -- This is an example of a raw data file.

U C L A CAMPUS COMPUTING NETWORK

18 / 35

ICPSR: Find Data

<http://www.icpsr.umich.edu/icpsrweb/ICPSR>

FIND DATA • START SHARING DATA • MEMBERSHIP • SUMMER PROGRAM • TEACHING & LEARNING • DATA MANAGEMENT & CURATION

ICPSR

Find & Analyze Data

Log In/Create Account

FIND DATA • SEARCH/COMPARE VARIABLES • DATA-RELATED PUBLICATIONS • RESOURCES FOR STUDENTS • HELP

Find Data

More Search Options

- Browse by topic
- Browse by sector
- Browse by geography
- Browse by investigator
- Browse by thematic collection
- Browse the national data
- View all studies
 - View all studies for which online analysis is available
 - View all replication datasets (more info)
 - View all studies that have teaching modules

Restricted-Use Data

ICPSR ensures respondent confidentiality within public use versions of the datasets. The vast majority of ICPSR data holdings are public-use files with no access restrictions. Sometimes the protective measures taken to reduce disclosure risk would significantly degrade the research potential of the data. In these cases, ICPSR provides access to restricted use versions that retain confidential data by imposing stringent requirements for accessing them. [Learn More](#)

Search Tips

Select a link below to reveal search tips relevant to your search:

I'm looking for datasets related to a specific topic.

I'm looking for datasets that contain specific variables.

I'm looking for datasets that answer a particular research question.

I'm looking for a specific dataset; I know the title and/or principal investigator.

f t v u G+ in

19 / 35

ICPSR: Register to Access Data

(ORCID Identifier: <https://orcid.org/>)

Log into ICPSR

Some functions of the web site, such as downloading and uploading data, require you to log in.

Login via Facebook/Google/LinkedIn

Sign in with

Google

LinkedIn

Facebook

Login via ORCID

Sign in Using ORCID

ORCID

ORCID[®] provides a persistent digital identifier that distinguishes researchers and supports automated linkages between their professional activities ensuring that their work is recognized.

Login via MyData

As an alternative to Facebook, Google, LinkedIn and ORCID, you can create a MyData account. Creating a MyData account is free and does not require you to be associated with an ICPSR member institution.

Returning User

Email

Password

Log In

New User

Creating a free MyData account will permit you to access features such as notification when data that interest you become available. Further, if you are from a [developing country](#), creating an account permits you to download data available only to ICPSR members.

Create Account

Request your password be sent to you via email.

Note: You must use your browser to access certain features in order to log in successfully. Your login session will last for 120 hours (or until you log out), after which you will be asked to log in again.

20 / 35

ICPSR: Search by Topics

Browse by Topic

Users have two options for identifying studies related to a particular topic:

- **Topic Classifications** address topics broadly, and are meant to be a good starting point for exploring the collection.
- The **Thesaurus** contains both broad and narrow terms that are more specific to the subject matter of individual studies.
- Our **Thematic Collections** are full websites that focus on particular subject areas.

Topic Classifications

I. Census Enumerations: Historical and Contemporary Population Characteristics

A. United States

1. Decennial Censuses

a. 1790-1960 Censuses

b. 1970 Census

c. 1980 Census

d. 1990 Census

e. 2000 Census

f. 2010 Census

2. American Housing Survey Series

3. Current Population Survey Series

4. Other Census, Including County and City Data Books

B. Nations Other Than the United States

II. Community and Urban Studies

A. Studies of Local Politics

1. United States

2. Nations Other Than the United States

D. Detroit Area Studies

C. Historical Urban and Community Structure Studies

III. Conflict, Aggression, Violence, War

A. Conflict and Stability Within Nations

B. Conflict Between and Among Nations

IV. Economic Behavior and Attitudes

d. Production Sector of Postwar American Economy

21 / 35

ICPSR: Search for and Compare Variables

Search/Compare Variables

The Social Science Variables Database (SSVD) enables ICPSR users to examine and compare variables and questions across studies or series. The SSVD currently includes over 4 million variables, representing about 70% of ICPSR's holdings that have quantitative data described in statistical syntax.

The SSVD does not include our entire collection. For a comprehensive search of all datasets, use our [main data search](#), which indexes all of the documentation available for our studies, including the variable descriptions.

college major

- [Browse by Study](#)
- [Browse by Series](#)

Comparing Variables

Compare variables using the checkboxes and the "Compare" button on the search results page.

Compare	Name	Label/Question
<input type="checkbox"/>	HERSON	dist of chelocherapil
		D3. Indicate the drug of choice (select all)
		Taken from: Criminal Justice Drug Abuse II
<input type="checkbox"/>	HERSON	USED HERSON AS MAIN DRUG
		Taken from: Services Research Outcomes II
<input type="checkbox"/>	HERSON	HERSON - ARKON LONG FIRST TIME FINE

Search Tips

- Enter words or strings that are likely to appear in a variable name, label, question, and value labels:

Presidential election will return variables dealing with all presidential elections.

Presidential election Obama will return only variables dealing with the 2008 and 2012 presidential elections.

- Use quotes to search for specific phrases:

"life satisfaction"
"minority rights"
"community progress"

- The minus sign may be used to remove certain types of results:

"Presidential debate" - "Bill Clinton" will eliminate the debates from 1962 and 1996 in which Clinton participated.

- A Boolean "and" is implied in the search.

- The search automatically does stemming (*studies* - *study*).¹ There's no need to type in an asterisk. It's also

22 / 35

Compare Variables

23 / 35

24 / 35

ICPSR: Reading About Data

[FIND DATA](#) [SEARCH/COMPARE VARIABLES](#) [DATA-RELATED PUBLICATIONS](#) [RESOURCES FOR STUDENTS](#) [HELP](#)

About the Bibliography

The *ICPSR Bibliography of Data-related Literature* is a searchable database that as of 2016 contains over 70,000 citations of published and unpublished works resulting from analyses of data held in the ICPSR archive. Developed with support from the National Science Foundation (SES-9977984), the *Bibliography* represents over 50 years of scholarship in the

Data-Related Publications

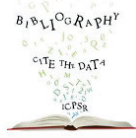
The *ICPSR Bibliography of Data-related Literature* is a continuously updated database of thousands of citations of [works using data held in the ICPSR archive](#). The works include journal articles, books, book chapters, government and agency reports, working papers, dissertations, conference papers, meeting presentations, unpublished manuscripts, magazine and newspaper articles, and audiovisual materials.

Please take a short survey [Ⓐ] to help us improve our data-related literature resources. ✕

- [View all citations](#)
- [Browse by author](#)
- [Browse by journal](#)
- [Browse by study](#)

Search Tips

• Our citation search includes only the literal citations, not the full text



25 / 35

ICPSR: Publications linked to Data

[FIND DATA](#) [SEARCH/COMPARE VARIABLES](#) [DATA-RELATED PUBLICATIONS](#) [RESOURCES FOR STUDENTS](#) [HELP](#)

Filter Results

Pub. Year ▾

Author ▾

Export Results ▾

Please take a short survey [Ⓐ] to help us improve our data-related literature resources. ✕

Search Results

Use the search box below to revise your search, or [start a new search](#).

[Studies](#) [Variables](#) [Publications \(1\)](#)

Your query returned 1 citations.

Sort by: Relevance ▾

	Pub. Year	Citation
1.	1992	Plomin, Robert, Scheier, Michael F., Bergeman, C.S., Pedersen, N.I., Nesselroade, J.R., McClearn, G.E. Optimism, pessimism and mental health: A twin/adoption analysis . <i>Personality and Individual Differences</i> . 13, (8), 921-930. Full Text Options: DOI [Ⓐ] Worldcat [Ⓐ] Google Scholar [Ⓐ] Export Options: BIS EndNote XML

Results 1 - 1

26 / 35

ICPSR: Publications Linked to Data

[FIND DATA](#) [SEARCH/COMPARE VARIABLES](#) [DATA-RELATED PUBLICATIONS](#) [RESOURCES FOR STUDENTS](#) [HELP](#)

Please take a [short survey](#) to help us improve our data-related literature resources.

Citation

The following citation is from the [ICPSR Bibliography of Data-related Literature](#):

[Plomin, Robert](#), [Scheier, Michael E.](#), [Bergeman, C.S.](#), [Pedersen, N.L.](#), [Nesselroade, J.B.](#), [McClearn, G.E.](#). Aug 1992. **Optimism and mental health: A twin/adoption analysis.** *Personality and Individual Differences*. 13, (8), 921-930.
doi:10.1016/0191-8869(92)90009-E

Full Text Options: [DOI](#) [WorldCat](#) [Google Scholar](#)

Export Citation: [BIS EndNote XML](#)

Related Studies

This publication is related to the following dataset(s):

- [Swedish Adoption/Twin Study on Aging \(SATSA\), 1984, 1987, 1990, 1993, 2004, 2007, and 2010](#)

27 / 35

Publishing and Data Management Plans

- What is a Data Management Plans
- Why do I need this?
- What does this have to do with Publishing?
- What tools are available to help?

28 / 35

What is a Data Management Plan?

A data management plan is a document that describes what you will do with your data *during* your research and *after* you complete your research.

29 / 35

Why do you need this?

- Funding Agencies
- Publishing
- Sharing and Archiving Data
- Regardless of the requirements, good data management is an essential skill for researchers.

Example: [Williams Institute](#), [UCLA School of Law](#) / [UCLA DataDen](#)

30 / 35

Basic Elements of a Data Management Plan

- What are your data?
- What formats will you be using?
- How will you describe this data?
- What intellectual property and privacy rights are associated with this data? (Will you need IRB approval?)
- How will you share this data? If you don't plan on sharing it, why not?
- How much will your data management cost?

See also [ICPSR Elements of a Data Management Plan](#)

31 / 35

Getting Help with Data Management Plans

- [UCLA Data Archive](#)
- [UCLA Library: Research Guides](#)
- [UKDA tools for creating and managing data \(Pdf document\)](#)
- [MANTRA – online learning tool/tutorial](#)
- [JISC Digital Media Advice](#)

ICPSR Resources for Data Management

- [ICPSR Framework for Creating a Data Management Plan](#)
- [Data Management and Curation](#)
- [ICPSR Guide to Archiving Social Science Data for Institutional Repositories, 5th edition](#)

32 / 35

Data Management and Confidentiality

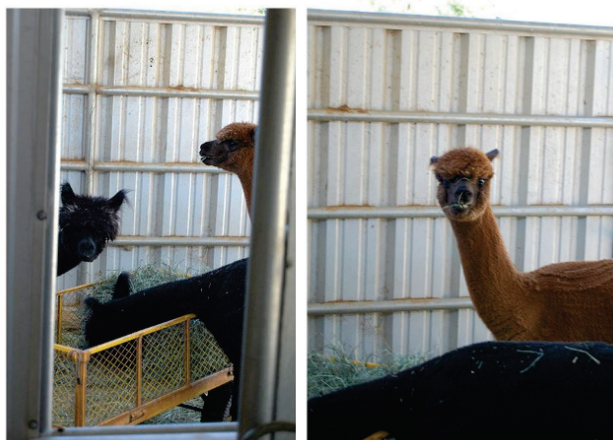
If your research involves human subjects, you will need to consider both legal and ethical obligations in managing and sharing your data.

Confidentiality refers to the agreement between the researcher and the participant about how the participant's identifiable private information will be handled, managed, and disseminated.

As a researcher, you need a clear view about how to protect the privacy of your research subjects.

33 / 35

Questions?



34 / 35

Problems ?

The Data Archive is here to help

1120-H Rolfe
ucla-data-archive@googlegroups.com
310-825-0716

Tim Dennis, Director of the Data Archive
Jamie Jamison, Archive Staff

[presentation](#)