

Research Data Management Basics

Faculty Development Day
January 24, 2014

Robin Camille Davis
Emerging Technologies & Distance Services Librarian
John Jay Library

Data curation

“The active and on-going management of data through its lifecycle of interest and usefulness to scholarly and educational activities across the sciences, social sciences, and the humanities.”

Source:

- guide.dhcurator.org/faq
- Craigin et al “An Educational Program on Data Curation” (2007)

Research Data Management

Resources and tools for managing your research data. (For faculty and graduate students.)

Last Updated: Oct 7, 2013 | URL: <http://guides.lib.jjay.cuny.edu/data-management> | [Print Guide](#) |

[About](#)[Data management plans](#)[Data security](#)[Storing & sharing data](#)[Metadata](#)[About](#)[Print Page](#)Search: [This Guide](#) ▾[Search](#)

Welcome!



This guide is intended for John Jay faculty and graduate students who work with research data. You might be new to data management or an old hand — either way, you will find useful resources here.

How and why do you 'manage data'?

Throughout your research project, you may be handling a lot of data: financial data, crime statistics, text corpora, biological data, etc. It is necessary that you understand how to stay organized and adhere to standards and best practices. Good data management leads to successful projects.

It will also be important for people outside of your project to be able to find and navigate your data in case you must provide further evidence in the form of raw data and if you decide to share the data as supplemental article material or archive it in a repository.

Furthermore, funders like the National Science Foundation (NSF) and the National Institutes of Health (NIH) require data management plans as part of grant applications. It is important for you to know what they expect, since part of your grant's chance of acceptance depends on a good data management plan.

Subject Specialist



Robin Davis
Emerging Tech Librarian

Contact Info
(212) 237-8261
[Send Email](#)

Links:
[Profile & Guides](#)

John Jay links

- [Office for the Advancement of Research \(OAR\)](#)
- [John Jay Library resources for faculty](#)
- [CUNY faculty research resources](#)

General data management resources

- [About data management + glossary](#)
The University of Minnesota explains what data management entails and provides a glossary of data management-related terms.
- [Guide to managing your data](#)
A subject guide from MIT that provides information, links, and other resources

Is this guide useful?

Was this information helpful?

☐ Yes ☐ No ☐ Don't Know

How useful is this page?
(1 = Not Useful, 5 = Very Useful!)

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Additional comments:

guides.lib.jjay.cuny.edu/data-management

A black and white photograph showing large stacks of bricks or blocks arranged in a semi-circular pattern, with a small table and chairs in the center, likely a hotel room where narcotics were found.

2013_12December.xlsx							
Search in Sheet							
Home Layout Tables Charts SmartArt Formulas Data Review							
Edit Font Alignment Number Format Cells Themes							
Paste Arial B I U % Conditional Formatting Styles Actions Themes Aa							
Align % % Conditional Formatting Styles Actions Themes Aa							
B107 = Fe Reference (quick; under 5 min.)							
	Timestamp	Reference type	Interaction type	Tags	Pattern type	Additional tags A	Time correction
103	12/20/2013 14:37	Technical			Grayscale		
104	12/20/2013 14:46	Reference; quick	In person	Research topic	Undergraduate		
105	12/20/2013 21:48	Reference; quick	In person	Research topic	Undergraduate		
106	12/20/2013 21:47	Reference; quick	In person	Research topic	Undergraduate		
107	12/20/2013 21:47	Reference; quick	In person	Research topic	Undergraduate		
108	12/20/2013 10:09	Reference; quick	In person	CLUNY, book search, printing	Undergraduate		
109	12/20/2013 18:51	Technical		Printing			
110	12/20/2013 18:58	Technical		CLUNY, book search, printing			
112	12/20/2013 10:02	Reference; quick					
113	12/20/2013 16:24	Reference; quick	In person				
114	12/20/2013 16:26	Reference; quick	In person				
115	12/20/2013 15:54	Reference; medium		Research topic	Undergraduate		
116	12/20/2013 11:01	Reference; medium		Research topic	Undergraduate		
117	12/20/2013 10:56	Reference; medium		Research topic	Undergraduate		
118	12/20/2013 11:27	Reference; medium	In person	Research topic	Undergraduate		
119	12/20/2013 11:27	Reference; medium	In person	Research topic	Undergraduate		
120	12/20/2013 11:34	Technical		Printing			
121	12/20/2013 11:34	Technical		Printing			
122	12/20/2013 12:08	Reference; medium	In person				
123	12/20/2013 12:01	Reference; medium	In person				
124	12/20/2013 12:07	Reference; quick	In person	Printing			
125	12/20/2013 12:07	Reference; quick	In person	Printing			
126	12/20/2013 12:18	Reference; medium					
127	12/20/2013 12:43	Reference; quick					
128	12/20/2013 12:43	Reference; quick		CLUNY, book search, printing			
129	12/20/2013 15:51	Technical	In person		Undergraduate		
130	12/20/2013 15:17	Directional	In person		Undergraduate		
131	12/20/2013 15:32	Technical	In person		Undergraduate		
132	12/20/2013 13:30	Reference; quick	In person	Research topic	Undergraduate		
133	12/20/2013 13:33	Reference; medium	In person		Undergraduate		
134	12/20/2013 13:33	Reference; medium	In person		Undergraduate		
135	12/20/2013 13:34	Reference; medium	In person		Undergraduate		
136	12/20/2013 13:39	Reference; quick	In person	Research topic	Undergraduate		
137	12/20/2013 13:39	Reference; quick	In person	CLUNY, book search, printing	Undergraduate		
138	12/20/2013 14:02	Reference; quick	In person	CLUNY, book search, printing	Undergraduate		
139	12/20/2013 14:21	Reference; quick	In person	CLUNY, book search, printing	Undergraduate		
140	12/20/2013 14:26	Reference; quick	In person	CLUNY, book search, printing	Undergraduate		
141	12/20/2013 14:26	Reference; quick	In person	CLUNY, book search, printing	Undergraduate		
142	12/20/2013 14:26	Reference; quick	Phone				
143	12/20/2013 15:52	Reference; quick	In person		Undergraduate		
144	12/20/2013 15:52	Reference; quick	In person		Undergraduate		
145	12/20/2013 14:02	Reference; quick	In person	Research topic	Undergraduate		
146	12/20/2013 14:04	Reference; quick	In person	Research topic	Undergraduate		
147	12/20/2013 14:09	Reference; medium	In person	CLUNY, book search, printing	Undergraduate		
148	12/20/2013 14:09	Reference; medium	In person	CLUNY, book search, printing	Undergraduate		
149	12/20/2013 14:18	Reference; quick	In person	Printing			
150	12/20/2013 14:18	Reference; quick	In person	Printing			
151	12/20/2013 14:21	Technical		CLUNY, book search, printing			
152	12/20/2013 14:41	Technical	In person	CLUNY, book search, printing	Undergraduate		
153	12/20/2013 14:41	Technical	In person	CLUNY, book search, printing	Undergraduate		
154	12/20/2013 14:41	Technical	In person	CLUNY, book search, printing	Undergraduate		
155	12/20/2013 14:41	Technical	In person	CLUNY, book search, printing	Undergraduate		
156	12/20/2013 14:41	Technical	In person	CLUNY, book search, printing	Undergraduate		
157	12/20/2013 14:41	Technical	In person	CLUNY, book search, printing	Undergraduate		
158	12/20/2013 14:41	Technical	In person	CLUNY, book search, printing	Undergraduate		
159	12/20/2013 14:41	Technical	In person	CLUNY, book search, printing	Undergraduate		
160	12/20/2013 14:41	Technical	In person	CLUNY, book search, printing	Undergraduate		
161	12/20/2013 14:41	Technical	In person	CLUNY, book search, printing	Undergraduate		

Lloyd Sealy Library Digital Collections, John Jay College of
Criminal Justice. <http://dc.lib.jjay.cuny.edu/>

branch master •

gephi-scripts / scopus-to-gephi_author-affiliation.py

robincamille 2 months ago Create scopus-to-gephi_author-affiliation.py

1 contributor

file 121 lines (93 sloz) 4.333 kb

Open Edit Raw Blame History Delete

```
1 # 2013-12-13
2 # Robin Camille Davis
3
4 # This script associates John Jay-affiliated authors from Scopus with the index keywords
5 # as assigned by Scopus. Authors and keywords are nodes with attributes Type (FacultyName, IndexKeyword)
6 # and Occurrence (how often each shows up).
7
8 # The data is (limited to science and social science publications, as indexed by Scopus.
9 # Scopus CSV should be structured as Authors,Year,Link,Affiliations,Authors with affiliations,Index Keywords
10 # or else you'll need to edit row counts.
11
12 from csv import reader
13
14 data = reader(open("scopus-auth-indexkeyword-affil-only.csv")) #file called Twice - see below
15
16 outfile = open("scopus-auth-indexkeyword-affil-2013y1118_2.gexf","w")
17 outfile.write('<xml version="1.0" encoding="UTF-8">\n')
18 gexf = open("http://www.gef.net/1.1draft/ version="1.1" xmlns:viz="http://www.gef.net/1.1draft/viz"
19 xmlns:xsl="http://www.w3.org/2001/XMLSchema-Instance" xsl:schemaLocation="http://www.gef.net/1.1draft
20 http://www.gef.net/1.1draft/gef.xsd">
21 <meta lastmodifieddate="2013-11-18">\n
22 <creator>robin.davis@creator>\n
23 <description>Github.com/robincamille/scopus-to-gephi-author-affiliation.py\n
24 <graph>\n
25 <attributes class="node" mode="static">\n
26 <attribute id="attr_type" type="type" mode="string"/></attribute>\n
27 <attribute id="global_occurrences" type="OccurrencesCount" type="Integer"/></attribute>\n
```

Code & software!

Tables & databases!

[illegible]

Text corpora!

Oral history interview with Ed Koch

DOCUMENT

PAGES

Zoom

Print

p. 1

Oral History Interview
with
Ed Koch

Interviewed by
Jeffrey A. Krosnick
February 29, 2008

Jeffrey A. Krosnick and Larry Sullivan
April 22, 2008

Justice in New York: An Oral History

No. 1

JOHN

Page 1 of 31

Description

Oral history interview with Ed Koch, Mayor of New York City from 1978-1989. Topics covered in the interview include the NYPD and its police commissioners, New York City's fiscal crisis, crime and race, capital punishment; crime in the subways; graffiti; Bernie Goetz and Sharpton.

Dates

February 29 2008 (Date interviewed)
April 22 2008 (Date interviewed)

Related collections

Justice in New York: An Oral History (is part of)

Related people

Jeffrey Krosnick (Interviewer)
Larry E Sullivan (Interviewer)
Ed Koch (Interviewee)
Ed Smith (mentioned in)
Bernie Goetz (mentioned in)
Al Sharpton (mentioned in)

Subjects

Koch, Ed, 1924-2013 – Interviews
New York (N.Y.) – Police Department
Crime – New York (State) – New York
Police – New York (State) – New York

Coverage

New York (N.Y.) – Social conditions

External Links

Audio clip: Koch on subway crime (1:56)
Audio clip: Koch on the 1977 campaign (3:19)
Audio clip: Koch on subway graffiti (4:42)
Audio clip: Koch on Al Sharpton (7:51)
Audio clip: Koch on Benjamin Ward (3:41)

Audio & transcripts! http://dc.lib.jjay.cuny.edu/index.php/Detail/Object/Show/object_id/1

Why care about data management?

- Good data management leads to successful projects
- Funders require good data management
- Sharing & reusing data depends on good data management from the get-go

Data management plans

- Summarize your project
- Who is responsible for the data?
- List the kinds of data your project will create and which formats you'll be using
- Identify the standards to which your data will adhere
- Define your plans for
 - data storage and security
 - sharing the data and policies of reuse
 - how you will archive and preserve the data



DATA CURATOR HULK

@DATACURATORHULK

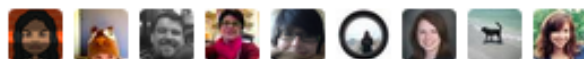
 Follow

YOU DOWN WITH D-M-P?
YEAH YOU KNOW ME
WHO'S DOWN WITH D-M-P?
EVERY LAST FUNDING AGENCY

 Reply  Retweet  Favorite  More

11

RETWEETS



8:22 AM - 26 Sep 2013

Example: Psychology (NSF)

DATA MANAGEMENT PLAN

Introduction

This Data Management Plan (DMP) covers the data which will be collected under the proposed project entitled 'Causal inference in visual identification: Integrating models, experimental paradigms, and behavioral measures' and is designed to be consistent with the NSF Policy on Dissemination and Sharing of Research Results. In accordance with this policy, this plan does not include preliminary analyses (including raw data), drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. Furthermore, data to enable peer review and publication/dissemination and/or to protect intellectual property may be temporarily withheld from distribution and other proposed data management. This plan will make certain that the data produced during the period of this project is appropriately managed to ensure its usability, access and preservation.

Description of Data

The types of data to be produced in the course of the project include background demographic information, accuracy and reaction time for the proposed visual identification tasks, and the results from computational model simulations. The demographic data is needed for published reports to convey the characteristics of the subject population.

Standards for Data and Metadata Format and Content

For preservation and long-term access, data collection will be accompanied with proper documentation and associated metadata. Files will include the data itself saved as Matlab (.mat) files, documentation files which a description of how the data was collected, and metadata in the form of the Matlab routines (Psychophysics toolbox) that administer each behavioral task.

Data Access, Sharing and Archiving

The researchers associated with this study are not aware of any reasons that might prohibit the sharing of the data to be generated under this project for public use and potential secondary uses. The principal investigators retain the right for first use of the data. The principal investigators will work with their NSF Program Officer to identify public databases which would be appropriate to house data generated from this project as well as for long-term archiving of the materials. Until an appropriate public database is identified, access to the de-identified data will be provided by contacting the PI. For journal publications, summary forms of the data (means and standard deviations) will be published in tables, appendices, or online supplementary materials. Raw data will in principle be available for access and sharing as soon as is reasonably possible, normally not longer than one year after publication of the data. The raw data from the proposed research will be archived locally by the researchers, on regularly backed-up computers and will be preserved for at least three years beyond the award period, as required by NSF guidelines.

Data Confidentiality

Research records will be kept confidential, and access will be limited to the PI and primary research team members. For each testing session, the recorded data will have any identifying information removed and will be relabeled with study code numbers. A database which relates study code numbers to consent forms and identifying information will be stored separately on password-protected computers in a secured, locked office. These computers are housed in research facilities in the Psychology Building at Indiana University-Bloomington, and in the Psychology Department at UCSD. A list of the names of individuals who have participated in each study will be maintained in order to ensure that no individual is tested more than once on related studies. To maintain the privacy of the participants, any report of individual data will only consist of performance measures without any demographic or identifying information.

Intellectual Property and Sharing of Research Resources

Intellectual property and data generated under this project will be administered in accordance with both University and NSF policies.

Ownership of sole or joint inventions developed under the project will be owned by the institution(s)

employing the inventor(s). Inventions shall be determined by U.S. Patent law, Title 35 SC. University and Participating investigators/institutions will disclose any inventions developed under the project and such inventions will be reported and managed as provided by NIH policies. Sole inventions will be administered by the institution employing the inventor. Joint inventions shall be administered based on mutual consultation between the parties. Similar procedures will be followed for copyrights.

Materials generated under the project will be disseminated in accordance with University/Participating institutional and NSF policies. Depending on such policies, materials may be transferred to others under the terms of a material transfer agreement.

Access to databases and associated software tools generated under the project will be available for educational, research and non-profit purposes. Such access will be provided using web-based applications, as appropriate.

Publication of data shall occur during the project, if appropriate, or at the end of the project, consistent with normal scientific practices. Research data that documents, supports and validates research findings will be made available after the main findings from the final research data set have been accepted for publication. Such research data will be redacted to prevent the disclosure of personal identifiers.

Example: GIS (NSF) p.1

This is an example Data Management Plan created by participants of a DataONE best Practices Workshop.

Data Management Plan

I. Types of Data

The products of this research are software to create maps, which are machine- and human-readable, containing information about the format and content of an HDF file. These maps can be used in the future to develop new software that enables reading of HDF files.

Map generation software will be developed for all platforms currently supported. Map generation software will be tested on the entire suite of HDF file types currently held by NASA. For quality control, individual archives that use the software to map their HDF data are responsible for verifying the quality of the maps created by performing a bit-by-bit comparison of data read using the maps and data read using existing HDF reading tools, presumably on a randomly chosen statistically significant subset of each type of HDF file in their holdings. To ensure the self-describing nature of the map files, non-HDF literate, non-native English speaking programming personnel will be used to develop the demonstration software described below. This will verify that the instruction and content of the map files are broadly understandable. The map files created as a result of this project will have the same name as the original HDF file from which they are derived, with the additional extension .map.

One ancillary data product produced by this project will be demonstration software (a reader) that uses the maps to read the HDF files, along with example input/output files and contextual information. While the existence of this demonstration software is not necessary to allow users in the future to read HDF data files, having a demonstration reader available in the near term will help the user community learn how to use the map files. The reader and associated files will be available as open source software at sourceForge.org, along with a copy held at the National Snow and Ice Data Center (NSIDC). Subversion will be used for source code and associated file version control.

Another ancillary product of this proposed work will be a collection of one of each kind of HDF file contained in the NASA archives. This collection of HDF files will be compiled and made available from the NSIDC archive.

The data manager during and after the project for the HDF file collection and the reader software will be Duerr. The HDF Group is responsible for management of the software during the project.

II. Data and Metadata Standards

The maps that will be produced are self-describing and require no documentation. Once created, the maps will become part of the data product, and will be included as one component of the metadata for the HDF product.

Example: GIS (NSF) p.2

III. Policies for Access and Sharing and Provisions for Appropriate Protection/Privacy

The user community for the map writer is expected to be any repository with HDF data. The user community for the map reader is expected to any user of HDF data in the future who wishes to re-use or adapt map-reading software to their own purposes. The user community for the collection of HDF files is expected be the community of tool developers who need a wide range of data types with which to test their product.

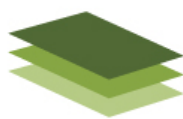
IV. Policies and Provisions for Re-Use, Re-Distribution

This proposed project will create ancillary metadata that will enable future users to read all HDF (Hierarchical Data Format) formatted data. The reader software is open source (licensed under GPLv3) and is freely available. The HDF data at NSIDC is freely available with no restrictions. The collection of HDF files has a suggested citation, which is described in the metadata.

V.

Plans for Archiving and Preservation of Access

The writer software will be preserved by the HDF Group for the life of the HDF libraries. The HDF Group uses industry-standard best practices to ensure the integrity of their software and systems. Once the map writer has been used to generate maps for every HDF file in existence, the continued existence of the writer software is not required. The reader software will be preserved at SourceForge.org for as long as there is community interest. The collection of HDF files will be preserved at NSIDC as long as utility is deemed high.



NSF-EHR: Education and Human Resources: 2. Period of data retention

EHR is committed to timely and rapid data distribution. However, it recognizes that types of data can vary widely and that acceptable norms also vary by scientific discipline. It is strongly committed, however, to the underlying principle of timely access, and applicants should address how this will be met in their Data Management Plan.

Progress

Click on a section below to edit it at any time.

✔ = complete

Plan description

1. Data generated by the project
2. **Period of data retention**
3. Data format and dissemination
4. Data storage and preservation of access
5. Additional possible data management requirements

Help

box size: [small](#) | [medium](#) | [full](#)

This section will allow you to account for any delay in the accessibility of your data after your research is done. Consider any reasons why you would not make the data immediately available —for instance, maybe you have political, commercial or patent concerns that will require you to postpone access to the data you produce. *Consider these questions:*

- How long will the original data collector/creator/principal investigator retain the right to use the data before opening it up to wider use?

[Preview](#)[Save](#)[Save and next](#)

Resources

General

[NSF Data Sharing Policy](#)

[NSF Data Management Plan Requirements](#)

[ICPSR – Effective Data Management](#)

[NSF EHR Directorate-wide Guidance](#)

[DataONE Best Practice: Sharing Data: Legal and Policy Considerations](#)

Metadata!

- “data about data”
- constructed information
- developed by people for a purpose

Source:

Coyle, K. “Understanding Metadata and Its Purpose.” *The Journal of Academic Librarianship*, Volume 31, Issue 2, March 2005, pp. 160–163.

[dx.doi.org/10.1016/j.acalib.2004.12.010](https://doi.org/10.1016/j.acalib.2004.12.010)

Metadata

- **Disciplinary standards**
 - Ecological Metadata Language
 - International virtual observatory specifications
 - Text Encoding Initiative
- **Discovery standards**
 - Title? Author? Date?


Data storage

(during research)


- Lots Of Copies Keep Stuff Safe (LOCKSS)
- Automate & test your backup system

Data security

[www.jjay.cuny.edu/
academics/1360.php](http://www.jjay.cuny.edu/academics/1360.php)



INFORMATION FOR:
Prospective Students
Current Students
Parents and Families
Faculty & Staff
Alumni

 **Web Apps**
Calendar / CUNY Services
JStop / Inside JJC / Work at John Jay
WebMail / Blackboard / Directory

Search Go

HOME | ABOUT US | ADMISSIONS | ACADEMICS | LIBRARY | STUDENT AFFAIRS | RESEARCH | NEWSROOM | GIVE TO JOHN JAY

Computer Security @ John Jay

Computer Security @ John Jay

**CUNYfirst /
NETWORK STATUS**

The DoIT - Technology
HelpDesk is located in
room L2.73.00 NB.

QUICKLINKS

- [Technology Helpdesk](#)
- [Inside John Jay FAQs](#)
- [Computer Security @ John Jay](#)
- [Right Answers @ John Jay](#)
- [Computer Equipment Request Form](#)
- [Audio Visual Request Form](#)
- [Password Reset](#)
- [Tech Bites](#)
- [Wireless Information](#)
- [DoIT Global Email Archive](#)
- [Site Licensing Information](#)
- [CUNY Portal](#)
- [Global Voicemail](#)

[Back](#)

A GUIDE TO PROTECTING YOUR COMPUTER & PERSONAL INFORMATION

WHAT YOU NEED TO KNOW?

The Department of Information Technology (DoIT) is responsible for safeguarding John Jay computers and networks from security breaches. We at DoIT are proactive in protecting the college's computing assets with regular systems upgrades, software patches, anti-virus software, etc. However, if you install spurious software or applications such as games, music or videos either from an untrusted source on the Internet or from infected media such as a flash drive or CD your computer can get infected and become compromised.

Hackers who gain control of on campus computers then use them to further attack other computers or the entire network system of the College. In most cases the focus of these hackers is to extract confidential data they can use for financial gain. It is important that as a technology user at John Jay you understand the responsibility you share with DoIT to protect against this ever present threat.

SECURITY STARTS WITH YOU! DO YOUR PART!

The most important component of good computer security for the John Jay Community is not a firewall or some network security device but your understanding of what risks exist and what you can do to safeguard yourself and others against them.

AWARENESS IS YOUR BEST DEFENSE! EDUCATE YOURSELF AND EDUCATE OTHERS!

Data security

- CUNY requires encryption of all confidential data
- Store non-public university data on a secure server (not your computer or flash drive)
- CUNY policies: security.cuny.edu

Archiving your data

3 years? 5 years? Forever?

Long-term storage:

- provide documentation
- prepare the data for long-term storage
- straighten out institutional/organizational support
- archiving solutions:
 - department or institution
 - trusted repository

Data repositories

- online vault for materials contributed by many people.
- its contents can be publicly available, restricted to certain people, or embargoed

Data repositories:

- [ICPSR: social science data](#)
- [National Archive of Criminal Justice Data](#)
- [Resource Center for Minority Data](#)
- [DataOne: earth data](#)
- [Digital history and classics studies](#)
List of repositories for historians. Includes archeology, epigraphy, manuscript studies.
- [Zenodo](#)
Universal repository, launched May 2013 by CERN
- [Mega-list from Databib](#)
Searchable catalog of data repositories.
- [Mega-list from UM](#)
Sorted by subject area.
- [Mega-list from DataCite](#)
This Google Drive spreadsheet lists out many repositories by subject, location, and deposit requirements.

ICPSR Find & Analyze Data

[Find Data](#) [Search/Compare Variables](#) [Find Publications](#) [Resources for Students](#) [Get Help](#)

Our Mission

ICPSR advances and expands social and behavioral research, acting as a global leader in data stewardship and providing rich data resources and responsive educational opportunities for present and future generations.

Announcements RSS

[Webinar: Resources for Health Research from ICPSR](#)

[China Data Center Event to Introduce Visiting Scholars](#)

[Project TIER to Host Workshop on Replication of Research](#)

[2012-2013 ICPSR Annual Report Released](#)

[Webinar on Social Science Data Management and Curation Features ICPSR Presenters](#)

[More...](#) ↗

[Subscribe to Email Updates](#)

Core Partners

ICPSR also distributes data provided by our partnering archives:

[Child Care and Early](#)

Find Data

More Search Options

- [Browse by topic](#)
- [Browse by series](#)
- [Browse by geography](#)
- [Browse by investigator](#)
- [Browse international data](#)
- [View all studies](#)
 - [View all studies for which online analysis is available](#)
 - [View all replication datasets \(more info\)](#)
 - [View all studies that have learning guides](#)

Restricted 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 restrictions on their access. 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. ([more](#))

Search Tips

Our search indexes the full documentation for datasets, including the variables.

Effective Searching

- You can enter a literal research question:
Do children of Asian immigrants speak English in the home more often than children of Latino immigrants?
- Or just keywords:
retirement "health insurance"
- To find studies with specific variables, separate the variables with commas and sort by "Variable Relevance" on the results page.

Search Syntax

- A Boolean "and" is inferred with multiple terms.
- [Stemming](#) ↗ is automatic; do not use an asterisk.
- Search is not case-sensitive.

icpsr.umich.edu

Sharing your data

- Reproducibility
- Funder requirements (e.g. NIH)
 - White House memo: research funded by major federal agencies must be publicly accessible
- For science! / For the humanities!
- Impact

From “Sharing Detailed Research Data Is Associated with Increased Citation Rate”

“Principal Findings:

- We examined the citation history of 85 cancer microarray clinical trial publications with respect to the availability of their data. ...
- Publicly available data was significantly ($p = 0.006$) associated with a 69% increase in citations, independently of journal impact factor, date of publication, and author country of origin using linear regression.”

Source:

Piwowar H., Day R., Fridsma D. (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLoS ONE* 2(3): e308. doi:10.1371/journal.pone.0000308

Thanks! Questions?

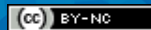
Robin Camille Davis

Emerging Technologies & Distance Services Librarian

John Jay Library

robdavis@jjay.cuny.edu « *not rdavis!*

@robincamille



This work is licensed under a Creative Commons
Attribution-NonCommercial 4.0 International License.