

# Data Librarianship: Past, Present, Future, Challenges, Opportunities

Stefan Kramer

Research Data Management Librarian

Cornell Institute for Social and  
Economic Research (CISER)

Presentation to GESIS, Bonn, Dec. 7, 2010



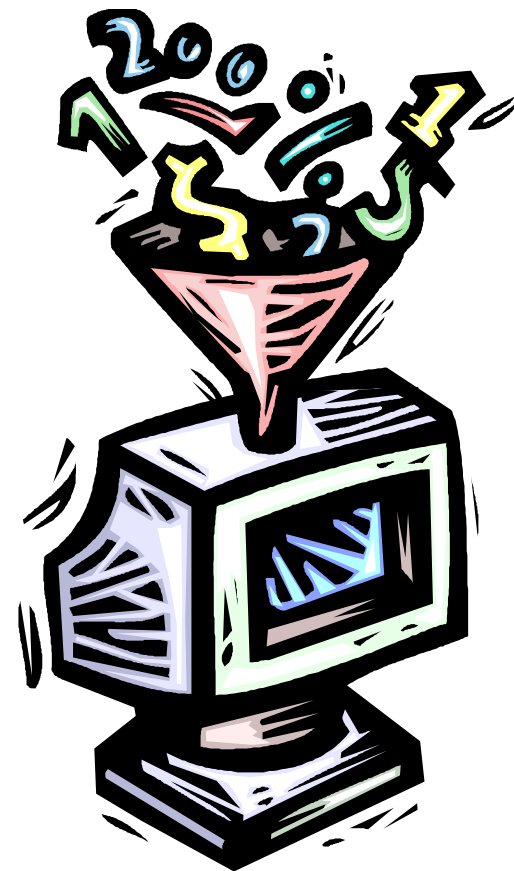
Cornell University  
Cornell Institute for Social and Economic Research

# What is CISER?

## What is CISER?

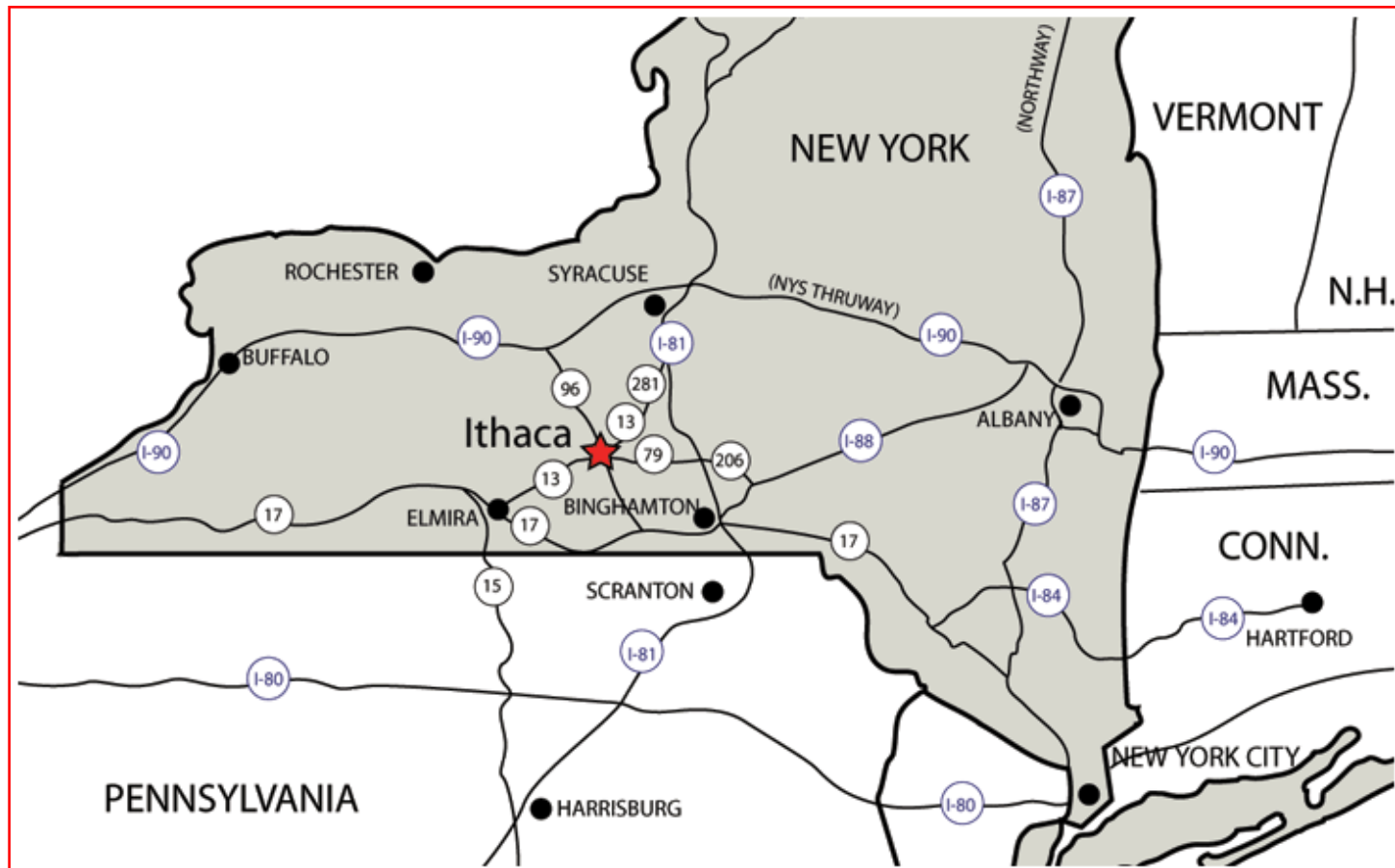
The Cornell Institute for Social and Economic Research was founded in 1981. Our mission is to anticipate and support the evolving computational and data needs of Cornell social scientists and economists throughout the entire research process and data life cycle.

More at: <http://ciser.cornell.edu/>



# Where is Cornell University?

Source: <http://www.cornell.edu/maps/state.cfm>



## Data librarians in German-speaking countries

- [Datenbibliothekare in Deutschland, Schweiz, Oesterreich?](#) (June 2009 posting to INETBIB list)
- *Diplomarbeit* by Stefanie Rümpel, FH Potsdam: [Data Librarianship – Anforderungen an Bibliothekare im Forschungsdatenmanagement](#)



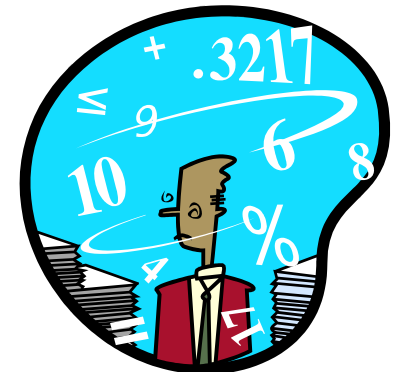
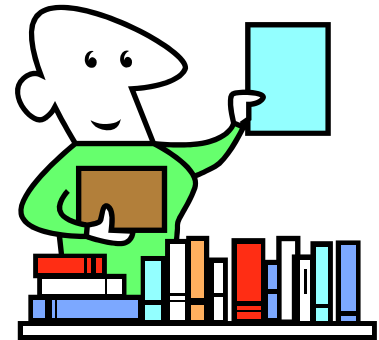
# Data librarians in English-speaking countries

- several dozen data librarians at larger research universities in Canada and the United States, a few in the United Kingdom
  - aside from positions at data archives that exist outside of libraries in universities (example: Cornell U.; UCLA)
- Majority in the social sciences, many of them members of [IASSIST](#)
- Most deal with traditional library activities and materials in addition to data, but really data-focused positions for librarians are emerging



# Data vs. other librarians in U.S. academic libraries

- Commonalities:
  - **most** academic data librarians also acquire journals, books, provide instruction, participate in general reference services
  - helping end users determine whether they need [statistics or data](#)
  - guiding and educating users on the information resources they need



# Data vs. other librarians in U.S. academic libraries

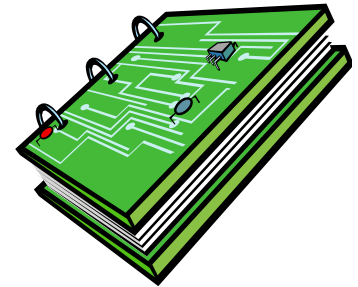
*Disclaimer: there also data librarians in non-academic institutions, e.g. RAND Corp., National Bureau of Econ. Research – not considered here*

- Differences: data librarians...
  - Need a certain understanding of quantitative research methods, and of file formats of (and compatible with) statistical software packages
  - Need to work closely with statistical (computing) consulting units on campus  
... examples: Yale's [StatLab](#), CISER's [Helpdesk](#)



# Data librarians also...

- Need to understand what documentation is required for data for it to be usable by researchers (such as codebook, questionnaire)
- For purchasing and licensing data, need to be able to communicate with vendors about how data can and must be accessed or downloaded, and by whom, and in what formats it will be available



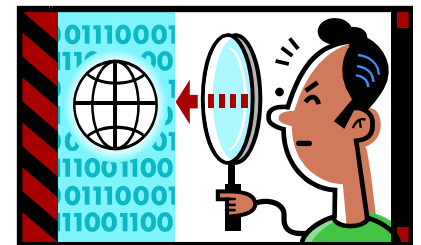
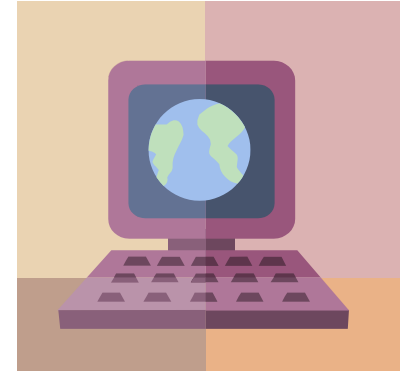
# Data librarians also...

- May need to repackage provider-delivered data and/or documentation for users – such as renaming, rearranging, manipulating, and compressing files
- May need to apply metadata that is specific to datasets, such as that developed by the [Data Documentation Initiative](#)



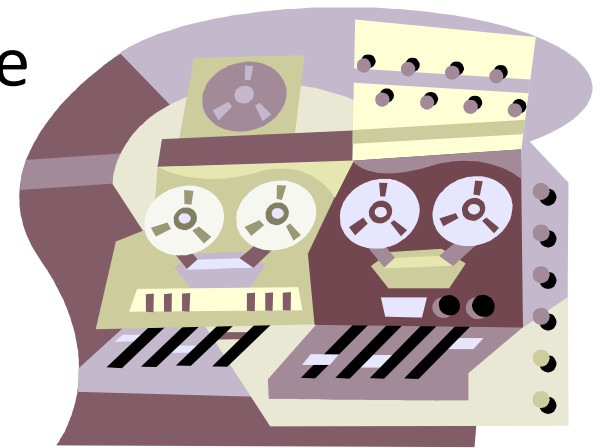
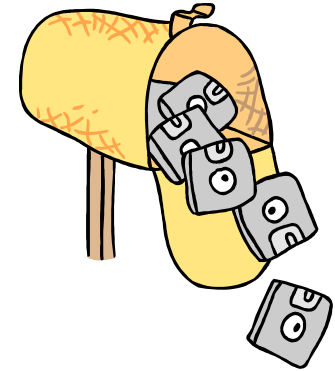
# Data librarians also...

- may need to know fairly little, or a lot, about Geographic Information Services, depending on separate map/GIS dept. or librarian for this being available
- need to understand the geographic *extent* and *granularity* of available data, because that is often an essential parameter of what users are looking for



# A brief history of data librarianship (esp. in USA & Canada)

- data librarians came into being at the intersection of typical IT and library activities
- From 1960s into 1980s, acquiring data for researchers usually meant ordering the data on magnetic tape from the data publisher, and mounting it on mainframes in computer center



# A brief history of data librarianship

- more technical than acquiring printed materials because data librarian has to understand the requirements of the researcher regarding data and file format, and the capabilities of the computing hardware available
- very library-like in that printed documentation to go with the data had to also be acquired, then organized and made available.



# A brief history of data librarianship

- In the 1990s, librarians played an increasing role in connecting researchers with data when the availability of statistical and GIS products from the 1990 *U.S. Census of Population and Housing* on CD-ROM made such data suddenly more accessible and of interest to a wider variety of users ... and on PCs



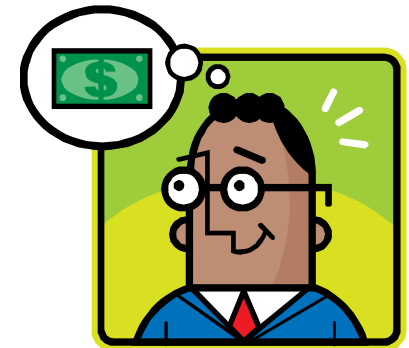
# A brief history of data librarianship

- Nowadays, statistical data and its documentation is very often made available to the library directly via the web – for example, through institutional membership in the [Roper Center for Public Opinion Research](#) or [ICPSR](#) – so the data librarian no longer necessarily needs to handle data or documentation locally
  - But there are exceptions, such as [CNTSDA](#)

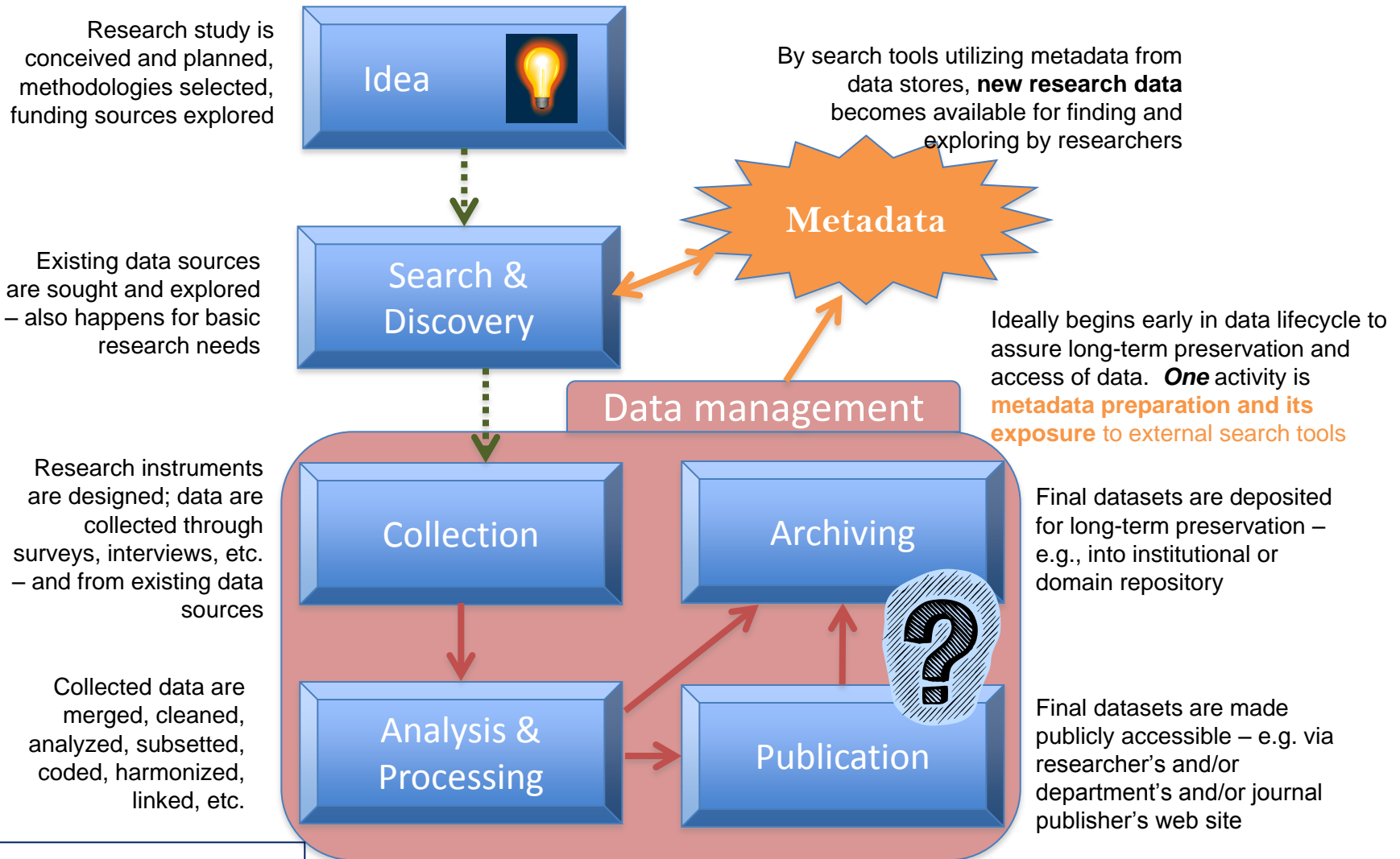


# The “next paradigm” in data librarianship

- potential significant future activity for data librarians: working with university faculty on managing and documenting ***their*** research data, sharing it as needed *during* the research process, and making it publicly accessible via a repository - be it the university’s own and/or a domain-specific one, such as [ICPSR’s Data Deposit program](#)
- major driver: forthcoming [National Science Foundation requirements for data management plans](#)



# Lifecycle of social science research data



From:  
<http://hdl.handle.net/1813/17472>



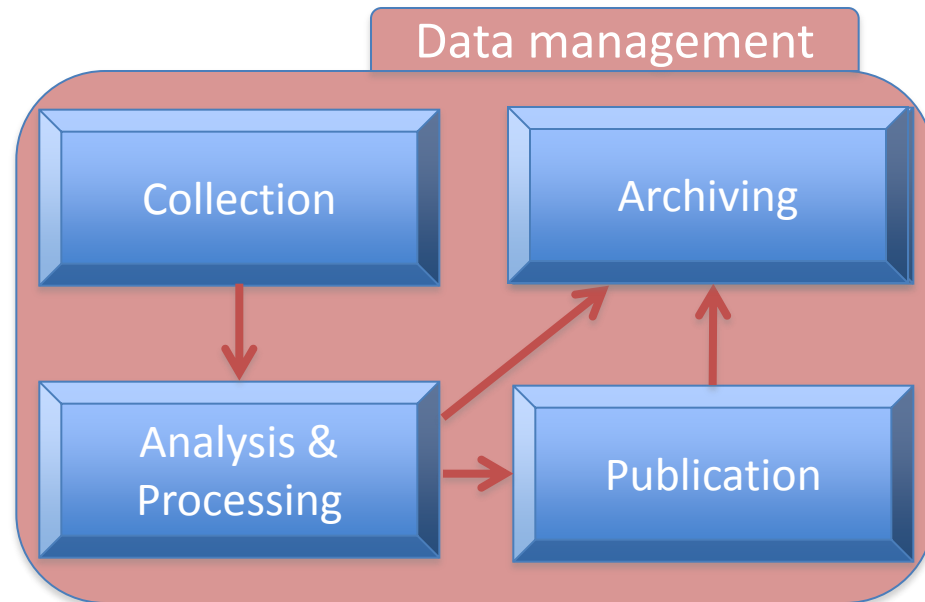
Cornell University  
Cornell Institute for Social and Economic Research

# Researcher buy-in is essential for data archiving

“Archives that preserve and disseminate social and behavioral data perform a critical service to the scholarly community and to society at large, ensuring that these culturally significant materials are accessible in perpetuity. **The success of the archiving endeavor, however, ultimately depends on researchers’ willingness to deposit their data and documentation for others to use.**”

ICPSR Guide to Social Science Data Preparation and Archiving: 4th Edition, p. 3  
<http://www.icpsr.umich.edu/files/ICPSR/access/dataprep.pdf>

Ideally, the archiving endeavor achieves researcher buy-in in **all** lifecycle stages involving data management activities – not just at the final point of archival deposit.



From:  
<http://hdl.handle.net/1813/17472>



Cornell University  
Cornell Institute for Social and Economic Research

# Making research data available for web-based analysis

- Most repository platforms make content, incl. datasets, available for *downloading*
- But for many audiences, such as introductory methodology classes or “the public,” analysis of downloaded data is asking too much (lacking software or skills)
- Possible solution: web-based analysis, exploration, visualization of *locally* created data, e.g. through [Berkeley SDA](#) or [Google Fusion Tables](#)

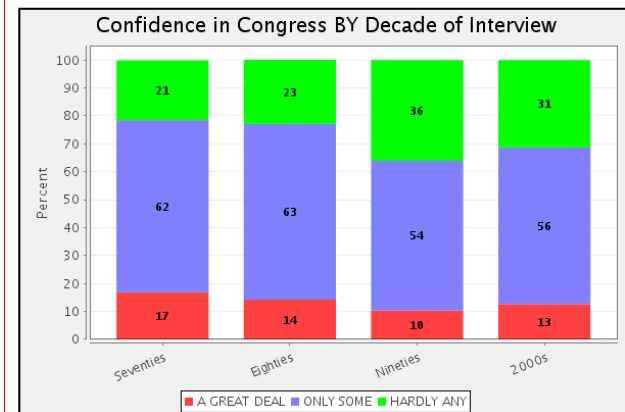


## Quick Tables

### Quick Table: GSS 1972-2008 Cumulative Datafile

Results: Confidence in Congress *BY Decade of Interview* (Percents)

	Seventies	Eighties	Nineties	2000s	TOTAL
A GREAT DEAL	16.8	14.2	10.3	12.6	13.6
ONLY SOME	61.7	62.9	53.6	56.1	59.0
HARDLY ANY	21.4	23.0	36.1	31.3	27.4
Total Percent	100.0	100.0	100.0	100.0	100.0
(Weighted N)	(8,770)	(10,893)	(8,507)	(6,817)	(34,988)
(Unweighted N)	(8,751)	(10,858)	(8,529)	(6,834)	(34,972)

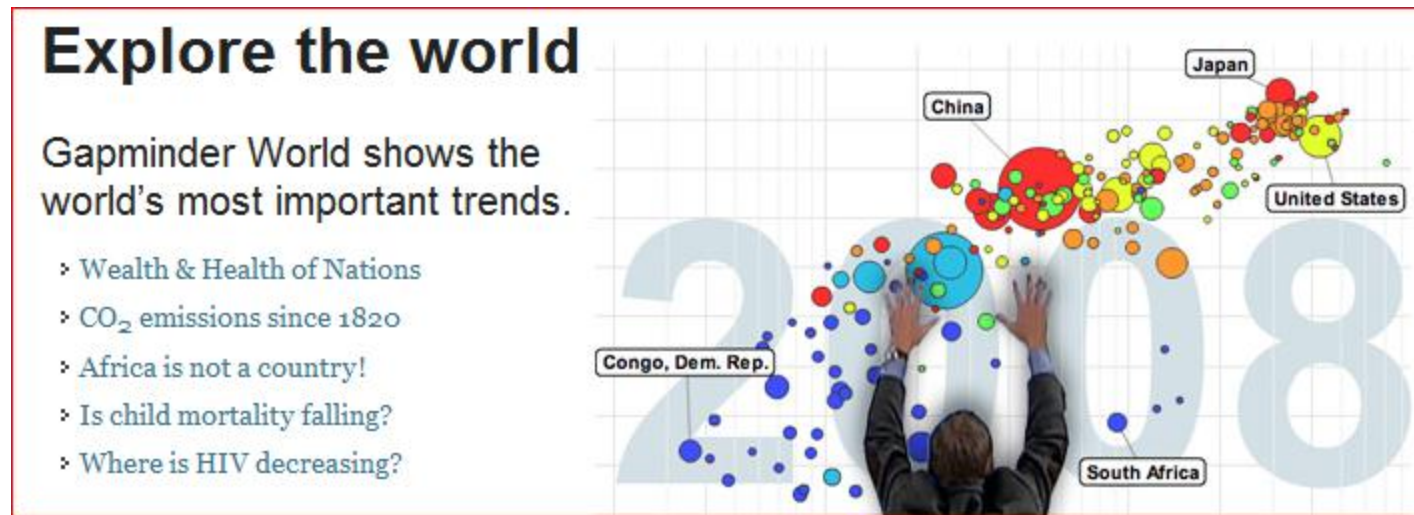


Output generated via <http://sda.berkeley.edu/quicktables/quickconfig.do?gss08>



# Making research data available for web-based *visualization*

- Could universities load locally created social science research data into a system like ... [Gapminder](#)?



# Thank you for your time & attention!



Stefan Kramer  
[stefan.kramer@cornell.edu](mailto:stefan.kramer@cornell.edu)



Cornell University  
Cornell Institute for Social and Economic Research