

# The Institute of Quantitative Social Science

Gary King





# The Social Science Data Revolution

- The Last 50 Years:
  - Survey research
  - Aggregate government statistics
  - In depth studies of individual places, people, or events
- The Next 50 Years: Spectacular changes due to . . .
  - Much more of the above
  - Shrinking computers & the growing Internet
  - Analogue-to-digital transformation of government records
  - The replication movement: academic data sharing (e.g., Dataverse)
  - Advances in methods
  - The march of quantification: through academia, professions, government, & commerce (*MoneyBall*, *SuperCrunchers*, *The Numerati*)
  - The end of the quantitative-qualitative divide:
    - Diverse perspectives remain but, with text, video, field notes etc. as data,
    - We now all benefit from the same technologies
    - Advantages of collaboration are growing

Social scientists need vastly improved infrastructure.

Harvard's response: IQSS -- a *hybrid organization*

- **A traditional research center:** facilitating research
- **Part of the Harvard administration:** building infrastructure
- **(Synergies:** turn administrative issues into quasi-research projects)

## IQSS Research Computing Environment (RCE):

- > 1300 shared computing cores
- 450 active users
- A persistent environment, sign on from anywhere
- Growing via collaborative research project (with Harvard CTO Jim Waldo and SEAS) to extend our platform to the Amazon cloud

- **Short courses:** R, Stata, SAS, Numeric Data Resources, Atlas.ti, GIS, & many IQSS technologies
- **Research consulting:** one-on-one help with data management and analysis
- **Survey Consulting:** help with question wording, sampling, analysis, nonstandard data collection
- Recently (re)grown with help from FAS

# OpenScholar: Faculty & Student Web sites



Personal Websites Exclusively for Harvard Scholars  
Not a Harvard Scholar? Find out how to create a similar site at your institution.

Scholars at Harvard offers you the next generation scholarly web site. With it, you can easily create and manage a highly dynamic, powerful, and customizable site, with an easy-to-use interface and a look and feel that is uniquely yours.

Fact Sheet (PDF) Video Intro

See who's using  
Scholars at Harvard!



Current Activity [\(more\)](#)

12 min 5 sec ago Venkatesh N. Murthy, Page updated: [Welcome to the Murthy Lab!](#)

© President & Fellows, Harvard University. Institute for Qu

Get Your Site Now



- > 800 **scholar** sites, including 300 by faculty
- > 120 **project** sites
- Free at scholar.harvard.edu  
(\$5-20K saved per site)
- Installation at > 50 other Universities
- Growing via collaboration (with HUIT and HPAC) to extend OpenScholar to **department** sites

# Dataverse Network



The Institute  
for Quantitative Social Science  
at Harvard University

Harvard's home for social science research.

Creating, sharing, and preserving scientific knowledge about human societies.

## IQSS Dataverse Network

[Search](#) [User Guides](#) [Report Issue](#)



### Welcome to the IQSS Dataverse Network!

Access the world's largest collection of social science research data across or browsing through one of the virtual data archives (called 'Dataverses'). You may also create a dataverse of your own, backed up in perpetuity in the **Murray Archive**, which may easily be customized to appear as if it were hosted on your own server. This means that the visibility, and access control for the data devolve to you, but all the guarantees, and software and hardware upgrades and maintenance are provided by IQSS. Learn more about the [Dataverse Network Project](#) ....

### Show All Dataverses

#### Filter by Organization

[Association, Society](#)  
[Business, Industry](#)  
[Government](#)  
[School \(Harvard\)](#) ▾  
[School \(Other\)](#)

#### Filter by Type

[Institutions, Depts, Centers](#)  
[Journal Replication Archive](#)  
[Large Collections](#)  
[Research Projects](#)  
[Scholars](#)

### All Dataverses

[ALL](#) [A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#)

Dataverses: **376** | Studies: **41,876** | Files: **671,000**

Name	View
<a href="#">Ansolabehere, Stephen</a>	<a href="#">View</a>
<a href="#">Leeper, Thomas</a>	<a href="#">View</a>
<a href="#">Harper, Sam</a>	<a href="#">View</a>
<a href="#">Druckman, James</a>	<a href="#">View</a>
<a href="#">Reingold, Beth</a>	<a href="#">View</a>
<a href="#">California Welfare Laboratory</a>	<a href="#">View</a>
<a href="#">Gries, Peter</a>	<a href="#">View</a>
<a href="#">Sorting Guilty Minds</a>	<a href="#">View</a>
<a href="#">Mikhaylov, Slava</a>	<a href="#">View</a>
<a href="#">Pettingill, Lindsay</a>	<a href="#">View</a>

- Your own **easy-to-use virtual archive**
- Our installation: ~ 1000 dataverses, 41,000 studies, 671,000 files
- > 15 installations in other Universities
- Growing via collaboration (with the Harvard Library) to expand infrastructure for diverse projects across humanities, social sciences, & sciences

- Easy, standardized solution to R's power and diversity
- $\approx 100$  statistical methods in common format
- $> 100,000$  users worldwide
- (Optional GUI through Dataverse, no R knowledge required)
- Large network effects: students learning from each other
- Growing via (open source) contributions



# Program on Automated Text Analysis

- “Computer-assisted reading”, a new way to organize and make sense of huge amounts of unstructured text.
- Advising with existing approaches available now
- New approaches under development, for public use in 2012

**Text Clustering**

Set Name: NSFC  
Size: 252 Files  
Description: NSF complete corpus, updatec

Clusters:   5 (Low)  15 (Medium)  30 (High)  
 Discoverable

Method:  Calc Mode:

Coordinates: -0.95312, 0.85156 Method: Mutual information Export Metadata  
Clusters: 5 Calc Mode: BASIC

Label	Percentage	Count	Detail
<a href="#">Label [+]</a> economists, markets, still, very, say, little, seem, price, quite, would	32.94%	83	<a href="#">Detail &gt;</a>
<a href="#">Label [+]</a> interdisciplinary, engage, integrated, building, within, scientific, create, capacity, collaboration, transformative	30.16%	76	<a href="#">Detail &gt;</a>
<a href="#">Label [+]</a> databases, record, access, digital, computer, technology, sources, technical, tools, capabilities	18.65%	47	<a href="#">Detail &gt;</a>
<a href="#">Label [+]</a> grand, indeed, obligation, noderivs, vice, gender, matching, opposite, throughout, partial	10.32%	26	<a href="#">Detail &gt;</a>
<a href="#">Label [+]</a> aimed, york, described, committee, does, similar, high, center, school, comparative	7.94%	20	<a href="#">Detail &gt;</a>

Display Method Points  Display History

Label	Coordinates	Clusters
-------	-------------	----------

- 37 active grants
- 30 – 40 proposals submitted annually
- Pre- and post-award support
- Working closely with department staff
- Social scientists from FAS and other schools *come* for help & *stay* to interact on research & teaching

# IQSS Activities Build a Community of Scholars

## **Administrative Services**

- Support for PIs, Programs and Centers
- Sponsored Research Administration

## **Technology Services**

- Research and Technology Consultants
- Technical Training
- Research Computing Environment (RCE)
- Repository for Social Science Data (Dataverse)
- Faculty Web Sites (OpenScholar)

## **Community & Collaboration**

## **Technology Development**

- OpenScholar
- Dataverse Network
- Zelig
- Text Clustering Tool
- Research Computing



## **Research/Affiliated Programs**

- Center for Geographic Analysis
- Program on Survey Research
- Program on Text Research
- RWJS in Health Policy Research
- Positive Political Economy
- Data Privacy Lab
- Workshops and Conferences
- Program on Methods