



**Share and Share Alike? Data-Sharing Practices in Different  
Disciplinary Domains**

**JoAnn Jacoby, University of Illinois at Urbana-Champaign**

**International Federation of Library Associations, Social Science Libraries  
Section, Satellite Conference**

**Disappearing disciplinary borders in the social science library - global studies or  
sea change?**

**University of Toronto, Toronto, Canada  
6-7 August 2008**

**Sponsored by:**

**CQ Press  
Worldwide Universities Network (WUN)  
Center for Global Studies (CGS) at the University of Illinois  
University of Illinois Library  
University of Toronto Libraries  
Faculty of Information, University of Toronto**

**This work is licensed under the Creative Commons  
Attribution-Noncommercial-Share Alike 3.0 United States License. To  
view a copy of this license, visit  
<http://creativecommons.org/licenses/by-nc-sa/3.0/us/> or send a letter  
to Creative Commons, 171 Second Street, Suite 300, San Francisco,  
California, 94105, USA.**

# Share and Share Alike? Data-Sharing Practices in Different Disciplinary Domains

JoAnn Jacoby

Associate Professor & Coordinator, New Service Model Programs

*Disappearing disciplinary borders in the social science library-  
global studies or sea change?*



*IFLA Social Science Libraries Section, Toronto, Canada, 6-7 August 2008*

[www.illinois.edu](http://www.illinois.edu)

# Outline

- When and how data is shared within a particular research community?
  - A model of factors shaping norms
  - Case studies: anthropology, economics & population studies
- Implications for interdisciplinary scholarship



# Why data-sharing?

- Funding agencies increasingly require that researchers deposit their raw data in appropriate public archives
  - facilitate the validation of results
  - enable secondary analysis & meta-analysis
  - ensure investment has greatest possible impact



# Why data-sharing?

- SHERPA-JULIET <[www.sherpa.ac.uk/juliet](http://www.sherpa.ac.uk/juliet)> research funders' open access policies, including data archiving
  - **encourage** data-sharing (NSF, OECD)
  - **require** data-sharing (Wellcome Trust, NIH grants >\$500,000)





Created with Wordle (© 2008 Jonathan Feinberg) <<http://wordle.net/>>, using much of the text from *To Share or not to Share: Publication and Quality Assurance of Research Data Outputs: Annex, June 2008* <<http://www.rin.ac.uk/data-publication>>.

# Investigating Data Curation Profiles Across Multiple Research Disciplines

IMLS funded project examining scholarly practice related to data:

“at which point in the research cycle are researchers willing to share data, with whom, and under what conditions?” <<http://cirss.lis.uiuc.edu/SciCom/cpp.html>>



Disappearing disciplinary borders in the social science library

# Social scientists share!

- Social sciences a leader in creating shared repositories & metadata standards
  - Many researchers routinely deposit data in large central repositories
    - ICPSR <http://www.icpsr.umich.edu/> was founded as consortium of political science data producers in 1962 and is now the world's largest archive of digital social science data.
    - United Kingdom Data Archive (UKDA), Council of European Social Science Data Archives





# Social scientists share!

## – Data Documentation Initiative (DDI)

- “An international effort to establish a standard for technical documentation describing social science data” <http://www.ddialliance.org/>
- DDI 3.0 was formally endorsed in May 2008



## ...And share alike?

- Repositories like ICPSR have not been adopted by all social scientists, nor are their solutions suited to the needs of all researchers.
  - Well equipped for numeric data (surveys)  
NOT other types of data
  - Selective acquisitions policy (political, economic, social survey data)



## ...And share alike?

- **Distributed, ad hoc data sharing**
  - Journals (*World Cultures*), Institutional Repositories, sites focusing on a particular research area (PaleoDB) or region (e.g., Portland regional development), researcher's own servers
- **No sharing at all**
  - Proprietary attitudes
  - Confidentiality
  - Logistical challenges



# Local example: Biological anthropologist

- Self-archives data on a local server and share it informally among a small circle of colleagues
- Now faced with NSF's requirement for a data sharing plan, he is exploring options for a more systematic approach
  - Biometric data – falls outside ICPSR's area of interest
  - Health-related repository deemed not suitable; not aligned with specific needs
- Older colleagues generally do *not* share

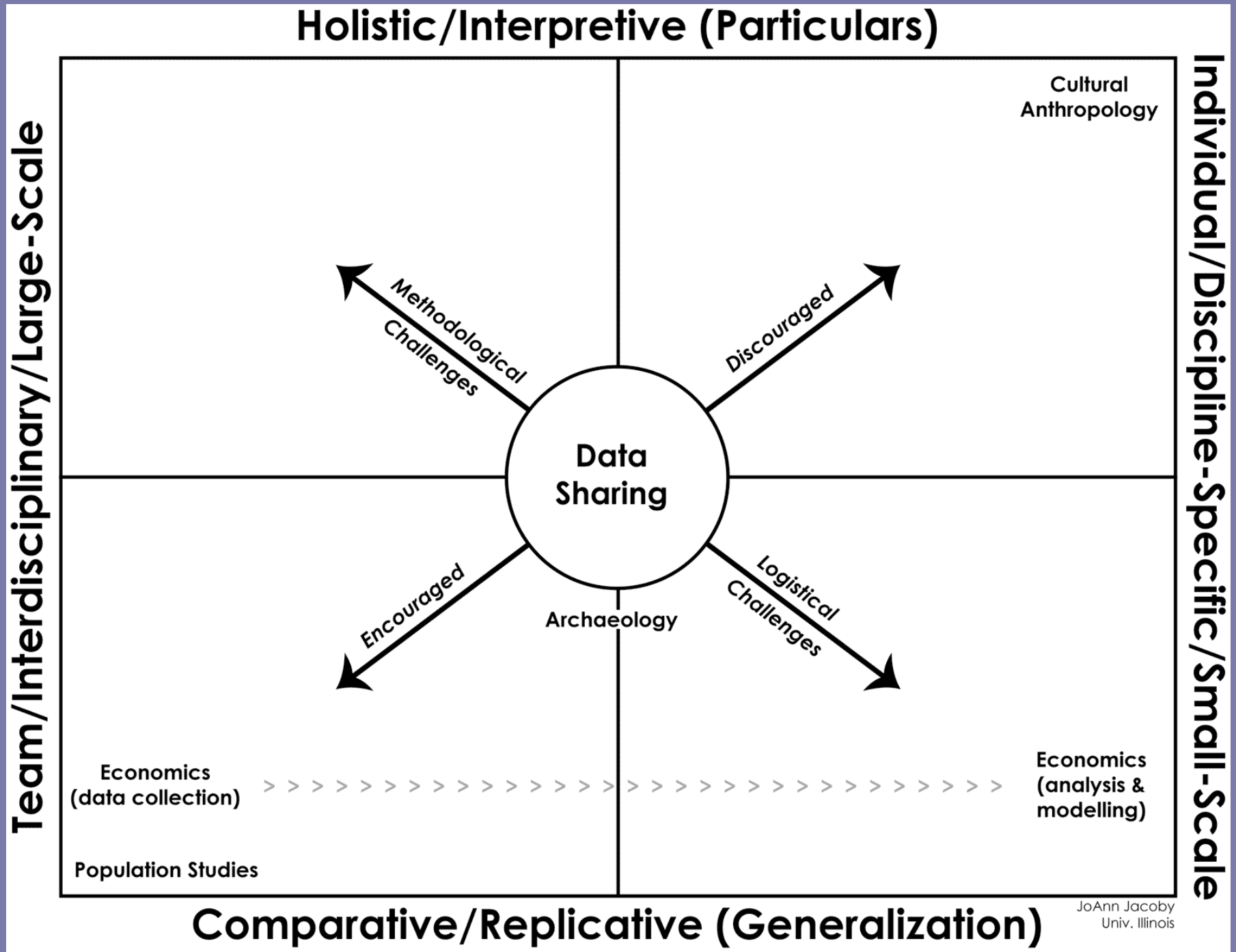


“There is not a perfect match between cultural norms ...and funder requirements. Some disciplines are well ahead of funding bodies in that they have had a culture of sharing data for a long time and have developed the infrastructures and methods for doing this. In other disciplines, data sharing is not commonplace and therefore funder policies may imply significant modifications to researchers’ attitudes and behaviour.”

(Research Information Network, 2008:12)



[www.illinois.edu](http://www.illinois.edu)



# What is social science data?

## Gutmann et al. 2004

### Survey data:

- Categorical or close-ended responses
- “Full-text” responses to open-ended questions

### Nonsurvey Data:

- Images
- Sound
- Video
- Multimedia, etc.

## Altman 2008

Raw measurements

Numeric tables

Administrative records (& email)

Video and audio interviews, transcripts (& blogs)

## Additional types of data

Fieldnotes  
(Silverman & Parezo 1995)

Texts

Models (e.g. *Economic Modelling* formerly ECONbase)

Artifacts, samples, other physical objects (Handler & Gould 1989)



# Social science data: Not just numbers

- Extremely heterogeneous
- Clearly, these “data” will not all fit in the same container (physical or metaphorical)!
- While physical objects may challenge our assumptions about what counts as data; these artifacts are familiar to museums, archives & special collections





# Case studies: Three Disciplinary Domains

- Anthropology
- Economics
- Population studies



# Anthropology

- Four subdisciplines (sociocultural, linguistic, archaeological and biological)
  - Holistic, interpretive → fieldwork
- Fieldnotes
  - Observations, censuses, genealogies, maps
- Texts, documents
- Objects, artifacts → images & spectrographs
- Samples, specimens



# Anthropology

- Proprietary attitude toward data
  - Context essential → holistic
    - Potential of misreading
  - Confidentiality / informant protection
    - Longterm, personal relationships
  - Laboriously gathered
    - Strong feeling of personal ownership



# Anthropology

- Data sharing increasingly common in archaeology and biological anthropology (but varies by research area & individual)
  - ALFREd
  - PaleoDB
  - NADB
    - Repatriation



# Economics

- Voracious users of data from local, regional and national governments
- Data from financial exchanges, banking manufacturing, prices, construction, etc.
- Corporate data would be of interest, but is not available in the public domain
  - Reliance on data gathered by others
- Models as data? (e.g. *Economic Modelling* formerly *ECONbase*)



# Economics

- Generate unique “micro-data”
  - **Micro-data** are disaggregated, broken down into finer level detail (e.g., expenses by household, imports of specific makes and models of cars)
- Interest in replication led journals to mandate complete datasets be accessible in the public domain, NSF & ICPSR partnerships to accomplish this starting in the late 1980s
- Before this, norm was *not* to share datasets created when performing secondary analysis
- Logistical challenges remain for small scale projects



# Population studies

- Study of fertility, migration, households (demography)
  - A “bridge discipline” that organizes the creative efforts of diverse scholars into large-scale data collection” (Evans 1991)\*\*
- Mixed methods (quantitative & qualitative)



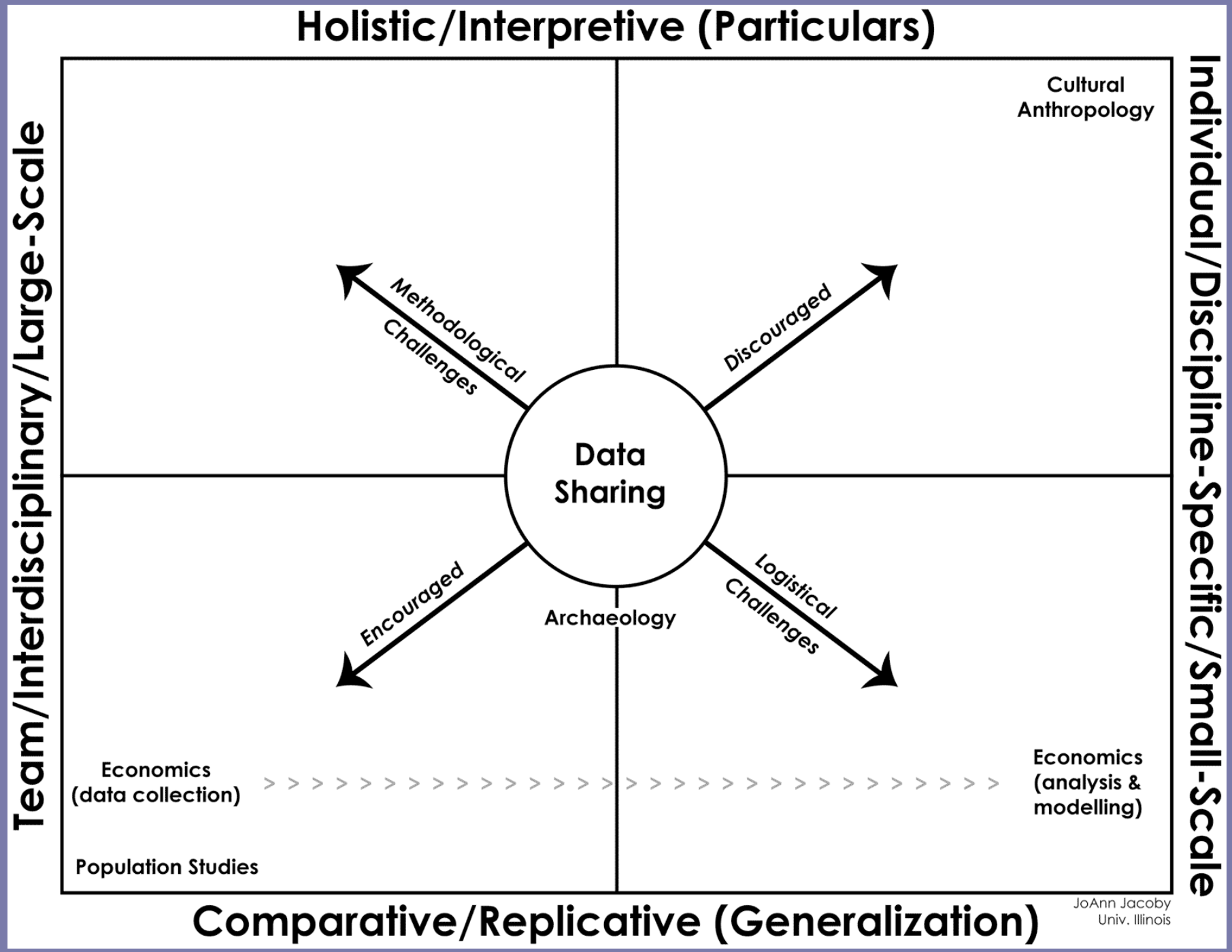
\*\*“Data Sharing in the Population Sciences” in *Sharing Social Science Data* (Joan E. Sieber, ed. 1991. Sage Publications)

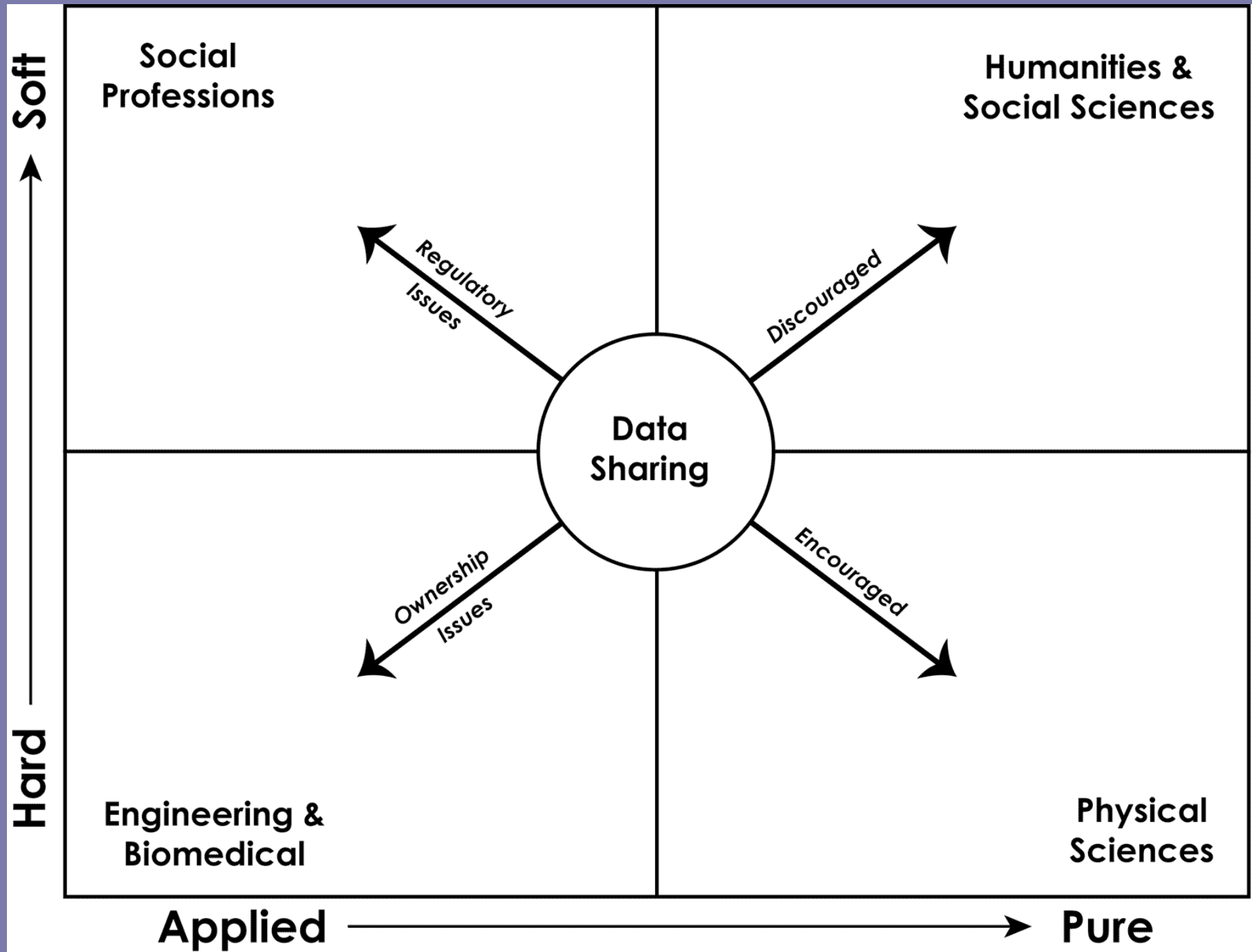
# Population studies

- Use government data gathered for administrative purpose (World Fertility Survey, GSS, census)
- May also be involved with data gathering & sampling
- Long tradition of data sharing
  - population-wide data gathering requires teamwork









JoAnn Jacoby Univ. Illinois

# Implications for interdisciplinary research

- Sharing data across disciplines with similar approach presents few challenges, except those of awareness & discovery
  - Quantitative sociology & political science  
(both well served by ICPSR)
- Or challenges inherent in the approach
  - Cultural anthropology & qualitative sociology, “data” may not be portable



# Data diversity

- Schröder (2007) cautions that “large scale data policies may have unplanned effects of homogeneity”
- For most of the history of science, data inextricable part of research process, useless when divorced from context
  - Mendel’s pea crossings!
- Need to encourage & foster data diversity



# Fear of miscegenation

- Variables from one data set recombined with others in “mash-ups” for meta-analysis (Myron Gutmann, IASSIST 2008)
- Methodology & rigor
- Who “owns” the new data set
- How to cite and track attribution at variable level?
- Provenance metadata, PURLs



# Sharing across domains

Balance the paradoxical need to “maintain data and documentation in a way that facilitates broad but appropriate use so that it continues to be useful to specific group of users, but can also be used by other, perhaps unanticipated user communities with very different needs and ways of approaching analysis.”

Parsons and Duerr (2005)



# Lost in the translation?

- Lakoff and Johnson (1980) on metaphor as basis for conceptual understanding
  - Metal: astronomy vs. common meaning
  - Race: demographer vs. anthropologist
- “These metaphors vary from discipline to discipline and are bound to change over time, even within a given discipline

(Parsons and Duerr 2005:32)



# Multidisciplinary, sure, but interdisciplinary?

Lack of a shared vocabulary are among the greatest challenges to interdisciplinary collaboration (according to the University of Minnesota study of the research behaviors of scientists & graduate students (Marcus et al. 2007))

*Perhaps also the deep semantic knowledge that underlies superficial differences in terminology?*





# The name of the rose

- Not only do the terms and concepts vary, but as we discussed previously, what counts as data.
- Even within communities of practice, differences in the metaphorical concept maps can cause dissonance (e.g., “Hispanic” label in the census, Bell 1996)



# Implicit knowledge

- Zimmerman (2003) dissertation:  
“Data sharing and secondary use of scientific data:  
Experiences of ecologists”
- Fieldwork shapes implicit knowledge →  
shapes reuse of data
- Involves formal & informal knowledge



# Informal and formal knowledge

- Importance of social exchange in scientific understanding
- Formal structures (standard research methods, metadata, storage formats) alone are ***not sufficient*** for large-scale data integration

Zimmerman 2003



# Making the implicit explicit

- Need to incorporate the informal aspects of gauging data quality
  - judgment, trust, shared understanding of the problem space
- Find ways to capture and communicate the implicit knowledge in ways that researchers recognize and can articulate

Zimmerman 2003



## Some researchers do cross the divide

- Anthropology (the “four fields” live under the same roof, but do not necessarily collaborate)
- Population Studies
- Health Sciences



# Librarians as envoys

- Accustomed to balancing needs of multiple audiences
- Generalist  $\leftrightarrow$  specialist divide
- Short-term access AND long-term stewardship



Thanks for your attention...

*Any questions?*  
*Thoughts?*

