Archives as a market regulator, or how can archives connect supply and demand?

Laurence Horton, Alexia Katsanidou
International Data Infrastructures
GESIS - Leibniz Institute for the Social Sciences
Data Archive for the Social Sciences
Abstract

- The data sharing movement
- Why researchers don’t want share
- What’s the point of archives?
- How our training center will SHOCK the world
The data sharing movement

“...a scientific e-infrastructure that supports seamless access, use, re-use, and trust of data. In a sense [...] the data themselves become the infrastructure – a valuable asset, on which science, technology, the economy and society can advance”

Riding the Wave, 2010.
“Your data is worth more if you give it away”

“Taxpayers have already paid for this information, the least we can do is give it back to those that want to use it new ways”

European Commissioner for Digital Agenda Neelie Kroes
United States of America

National Institute of Health, 2003
National Science Foundation, 2011
Institutional specific approaches
United Kingdom

Economic and Social Research Council, 2000
Research Councils UK, 2011
Federal Republic of Germany

German Research Council (DFG),
1998
Objections to sharing I

• Why me?
  – Surely my data isn’t interesting to anyone but me

• See it through my eyes
  – How can you reuse data for a different research question?
Objections to sharing II

- What’s in it for me?
  - Why should I miss out on the Nobel prize?
- Duty of care
  - I have to protect my participants.
The data management environment

- Mostly sticks and few carrots
- We need more carrots
The role of data archives

Data creators + Data archives = Data reuse
The role of data archives
Promoting reuse: cognition vs. emotion

“Educating the mind without educating the heart is no education at all”
Promoting reuse

Arguments for:
“cold” cognition
• Intellectually conscious
• Controlled
• Based on explicit learning

Arguments against:
“hot” emotion
• Instinctive
• Impulsive
• Reactive
Just DOI it…and other things

• Investing in tools to support data citation so researchers get credit
• Promoting data to get data discovered
• Promoting data repurposing
• Emphasizing and investing in archives as safe storage and dissemination services
The researcher’s greatest fear

The only thing we have fear is…

…our publications and data being ignored
The inevitable research data lifecycle slide

- Data preparation for reuse
- registration with DOI
- long-term preservation
- dissemination
- provision of tools & standards
- publishing research results through information systems

Research

- Reviewing existing data sources for potential re-use, replication, enhancement, and methodologies

Archiving and registering

- Working from copies not the raw data. Keep copies of syntax files and all analysis steps

Study planning

- Describe data to be collected: types, formats, standards, documentation, methods, ownership, consent, responsibilities, platforms for working storage, back-up, quality assurance, long-term preservation plans

Data collection

- Implementing methods of collection, description, assurance of data quality through checks and inspections
Activities

- workshops on data management and archiving topics:
  - data management planning
  - data sharing
  - best practice on consent for reuse, copyright and use of existing data, confidentiality and anonymization, documentation and data enhancement, methods of data sharing, file formats, physical and digital storage
  - long-term archiving and preservation, data dissemination and security, licensing data for use, managing access systems, format migration and verification
Our audiences

- **principal investigators** who plan data management and are responsible for oversight
- **researchers** who implement data management
- **individual researchers** to maximize the utility of data
- **data archivists** responsible for digital curation, enhancement, and long-term preservation
Manage your data

What does it mean to "manage your data?" It means that to keep data safe from harm and making data usable and accessible for a world where a value can survive is to pool it.

Thinking about and implementing research data management throughout the research life-cycle is essential to ensure data quality, confidentiality, preservation, access and reuse.

Responsibility for research data management lies not just with researchers but also with institutions and funders.

Research data management policies exist at institutional or national levels and significant investments have occurred in infrastructure to support research data management. However, this investment is often at an institutional level or is not discipline-specific. For example, see the Digital Curation Centre's Data Management Planning Online Tool.

The Archiving and Data Management Training and Information Centre is staffed by social scientists for social science concerns and to help you develop a research data management strategy. You can either contact us or attend one of our workshops.

- Data sharing
- Descriptive metadata
- File formats
- Data security
- Data backup and secure working storage and sharing during research
- Intellectual property rights
- Data consent and ethics
Looking after your research data: Introduction to research data management

Instructors: Dr. Alisca Katsanidou, Lawrence Hertan, Dr. Sergiu Ghunghina

June 27 (2pm) - June 28 (1pm), 2012

Looking after your research data is essential. Good data management practice minimizes the risk of data loss, ensures research integrity and facilitates replication; it enhances data security, research efficiency and reliability; and saves time and resources. Moreover, data management planning and data reuse is increasingly a requirement of funding organizations.

This workshop comprehensively discusses ways to ensure the security and integrity of your research data. It is designed to address the needs of social science researchers working with either qualitative or quantitative data.

Workshop sessions include basic concepts of data management, advice on writing a data management plan, licensing data for reuse or to reuse, consent and ethics for data reuse, file formats, data migration and metadata, data storage, back-up and security, data management in collaboration research, and archiving your data.

The target audience consists of three types of researcher:
1. Principal investigators who have to plan data management and ensure its implementation strengths.
2. Researchers who are members of project teams and for whom the implementation and maintenance of data management procedures is crucial.
3. Individual researchers who can maximize the utility of their data by adopting data management practices.

The workshop provides an interactive hands-on approach to looking after your research data and encourages discussion amongst participants on shared problems or experiences as well as the chance to consult with the International Data Infrastructures team at GESIS.

Learning objectives
DASISH

DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

About DASISH

Connecting Infrastructures

This project brings together all E-SCI (access to e-infrastructures) projects in the social sciences and humanities (SSH) to improve the cohesiveness of the infrastructure development and to work on a few common joint activities.

The rationale behind

The rationale behind this idea is that

1. all e-infrastructures should be promoted,
2. initiatives should mutually benefit from the advancement work of the others and
3. to establish joint integrated domains where this makes sense for the SSH users.

Joint activities will focus on the following dimensions: understanding the different architectural scenarios, assessing and improving data and metadata quality, setting up a tools and services forum, improving the quality of survey data, develop and improve data preservation and curation services, develop a joint shared data access and environment framework (SAA, PaaS), joint projects, Web of open data, and a joint coordination framework, jointly work on legal and ethical aspects, carry our much training and education efforts, work on disseminating the results.

DASISH is a FP7-INFRASTRUCTURES-2011-1 project

Grant Agreement 284649, Contribution of CF & DBA. The project duration is 36 months, starting on 1st January 2012 and ending on 31st December 2014.
Thank you

Archive Training Center
GESIS Leibniz Institute for Social Sciences
Unter Sachsenhausen 6-8
50667 Cologne
GERMANY

Email: archive.training@gesis.org
Twitter: @GesisID1