

Acquisition in a Growing Social Science Data Landscape Challenges and (Possible) Solutions

Oliver Watteler

(Data Archive for the Social Sciences)

CESSDA Expert Seminar 2012 (FSD), October, 4th – 5th

A Growing Social Science Data Landscape – Chance or Threat?

- Challenge: Not only growing amount of data relevant to researchers in the Social Sciences, but also growing number of data holding institutions (projects or facilities).
 - More data, but also more sources.
- Question: How does the user get to the data?

Structure of this presentation

- Focus is on the German situation.
- Details of the challenges in a growing Social Science data landscape.
- Short introduction to GESIS.
- Possible solutions to the challenges.

„Discovery“ of Archiving as a Topic

Challenges

- Topic of long-term preservation drew new interest to data sharing and data archiving (e.g. UNESCO 2003, National Academies 2009, Nestor 2009).
- Internet offers new channels of distribution!
- But, researchers still want their data to be saved!
- Advantage: Role of „archives“ is strengthened!
- Disadvantage: More stakeholders to talk to!

Growing Number of Institutions Offering Research Data

Challenge

- Growing number of institutions intend to offer research data!
- Advantage: Specialized knowledge, closer contact to researchers in particular fields!
- Disadvantage: Further fragmentation of data landscape!

Footnote: GESIS documents some 5800 addresses of research institutions in Germany, ~ 95% from the area of Social Sciences.

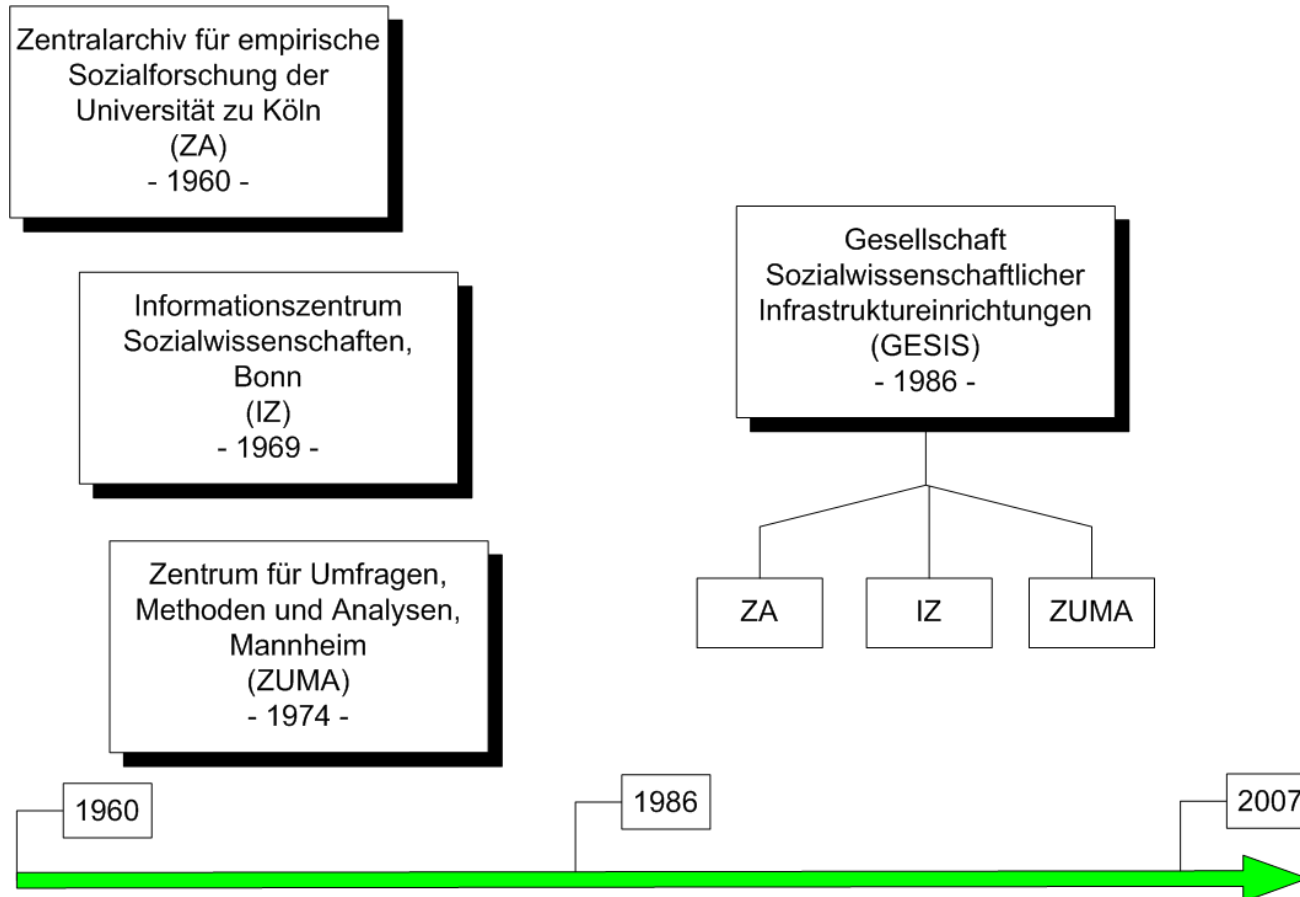
Herding cats: the German Data Forum

- Founded in 2001.
- Goals: Advance Social and Economic Data Infrastructure and improve access.
- (+) Encourages the founding of data holding facilities.
- (+) This way makes new data available.
- (−) It also contributes to fragmentation.
- 17 Research Data Centres (RDC) and 3 Data Service Centres (DSC).
- GESIS participates with four centres.

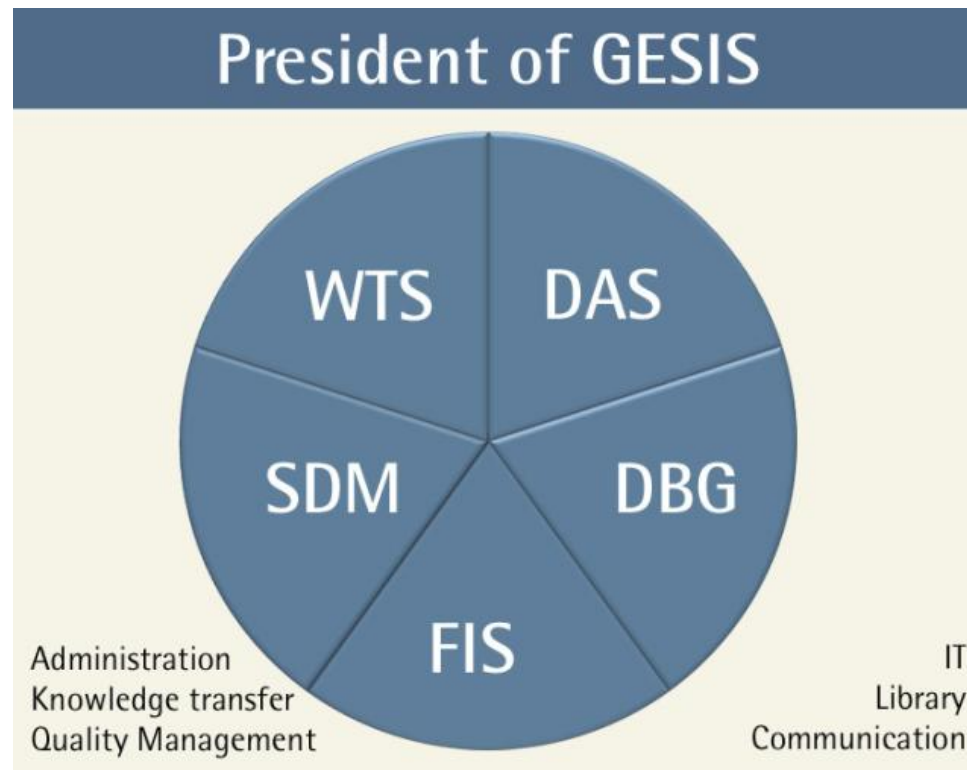
GESIS - the Institute

- Social Science Services since 1960.
- Basic, International and Research Oriented Services
- Staff: 300+
- Three Locations (Berlin, Cologne, Mannheim)
- Institutional development reflects changes in data landscape!

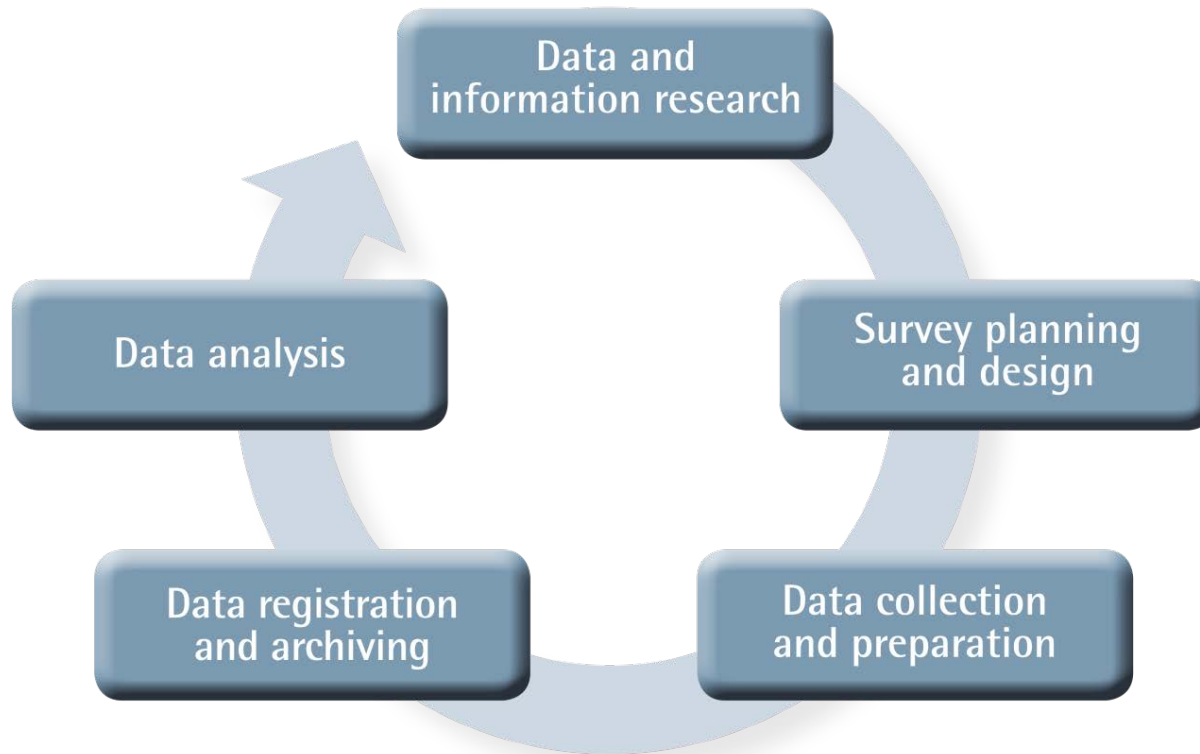
Origins and Development



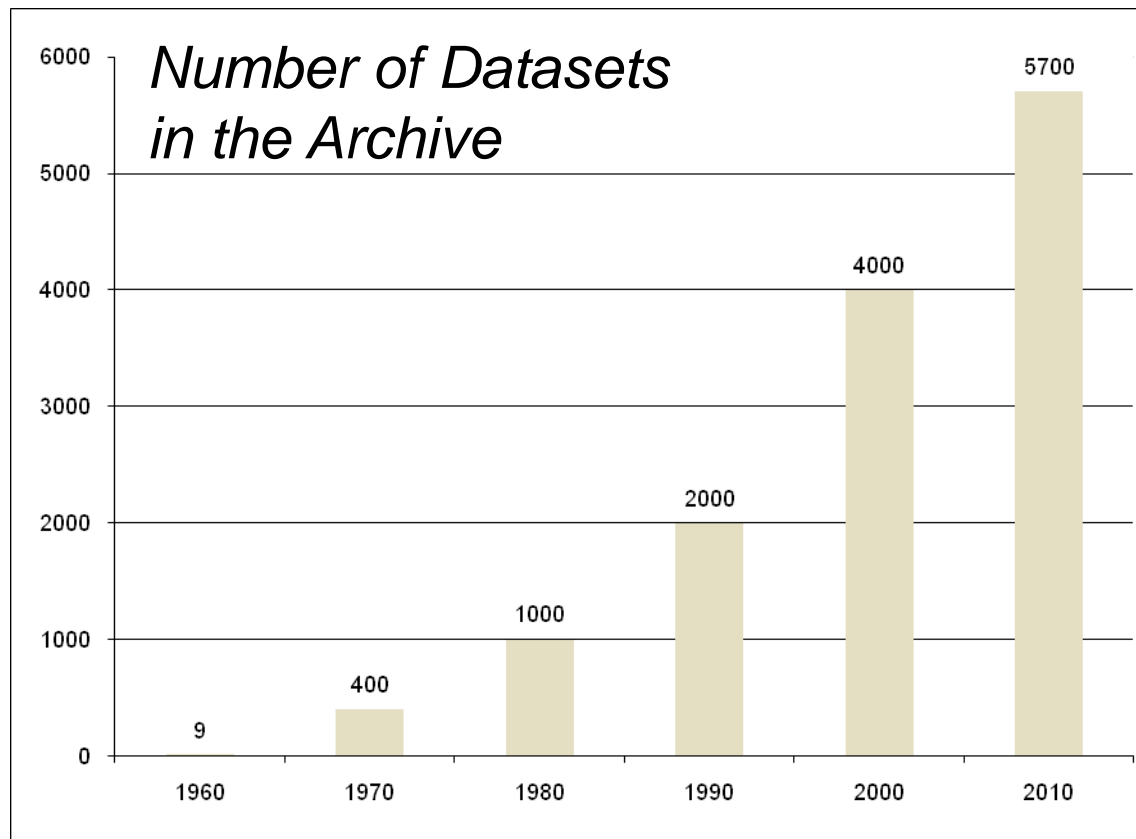
Status Quo - Departments



Services Along Life Cycle



Data Holdings (1)



Data Holdings (2)

- Survey Data
 - Various Topics
 - Cross-Section > Cross-Cultural over Time
 - Different Collection Modes
- Historical Data
 - Time Series Data
 - Data on Social Structure

Data Holdings (3)

- Comparative Study of Electoral Systems (with ICPSR)
- Eurobarometers (with ICPSR)
- European Values Study
- European Voter
- German National Election Studies since 1949
- International Social Survey Program
- ...

Acquisition then ...

- No obligation by funding agencies to archive.
- Researchers in the archive made their data available.
- Researchers established contacts or participated in projects. Data was deposited at the archive.
- Establishment of lasting relationships to researchers and institutions.
- The „few“ projects that produced data were contacted with a letter and printed information.

... and now.

- Still no obligation to archive by funding agencies, BUT strong recommendation.
- Close cooperation with researchers still important.
- Distinguish passive and active acquisition.
- Closer cooperation with data holding institutions.
- Set up of acquisition committee (similar to UKDA's Acquisition Review Committee).
- Offer of registration and documentation services.
- Other activities to support acquisition.

Focus of acquisition

- Passive and Active Acquisition:
 - Passive Acquisition in combination with „Self-Archiving“,
 - Active Acquisition of high quality data with an expected wide use for research community.

Cooperation with data holding institutions (1)



Cooperation with data holding institutions (2)

Example: Federal Institute for Vocational Education and Training (BIBB)

BIBB

- Original data at RDC (e.g. Employment Survey 2006)

GESIS

- Distribution of reduced scientific use file (SUF)
- Long-term Preservation of reduced SUF

Data Acquisition Committee

- Put together this year.
- Five Data Archive teams permanent members.
- Main aims:
 - Plan active acquisition activities!
 - Pull together information about ongoing research projects!
 - Prepare acquisition proposals for GESIS board if further resources are needed!
(E.g. funding for data preparation)
 - Report about acquisitions!

Some quality criteria for active acquisition (1)

- Data is suitable for answering research questions in the Social Sciences!
- International comparison is possible!
- Comparison over time is possible!
- Data is well prepared!
- Data is well documented!



Some quality criteria for active acquisition (2)

- Quantitative collection modes were applied!
- Data is of use for a wider audience in the research community!
- Data is methodologically or content wise unique!

Registration and documentation

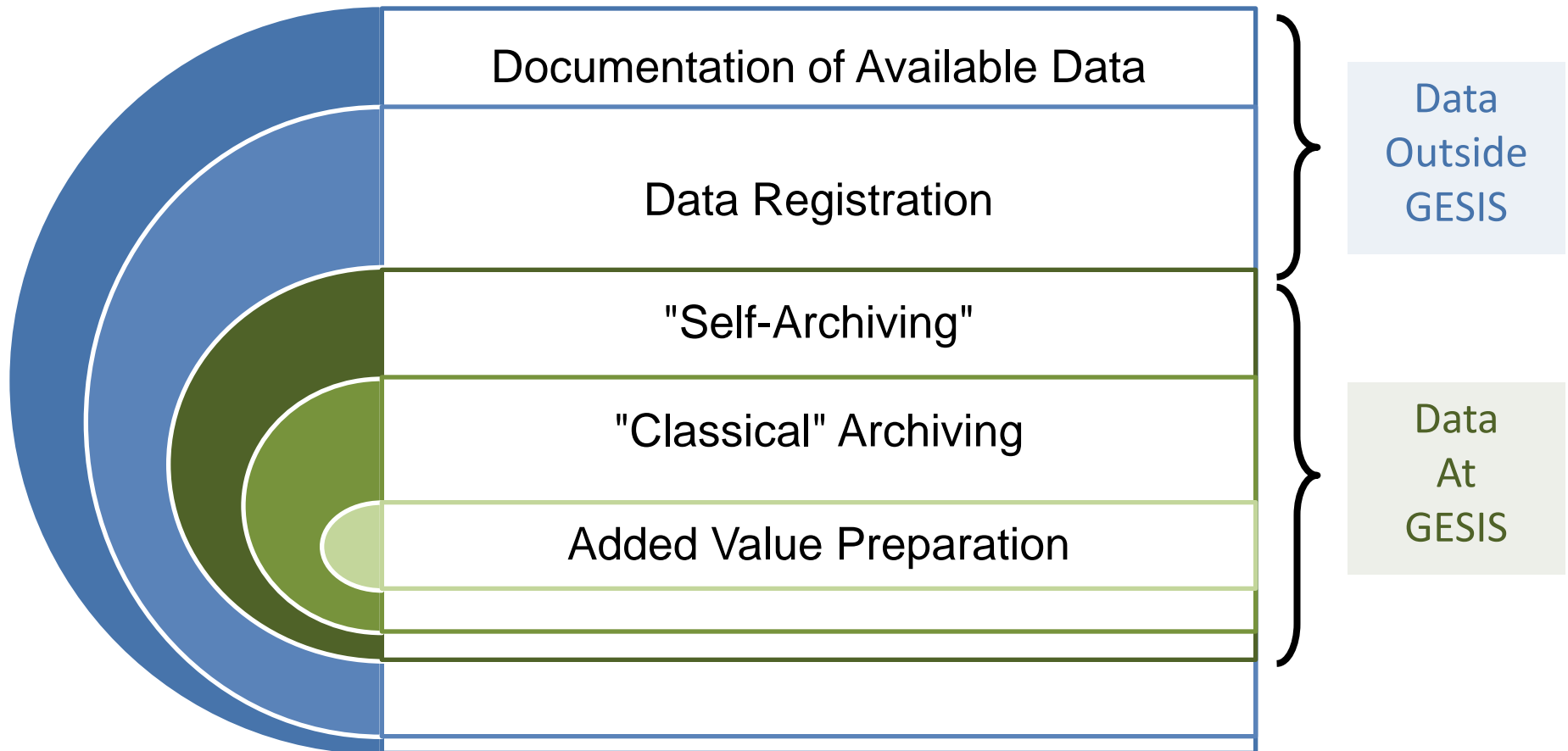
- **da|ra** Registration agency for social and economic data in cooperation with



- Additional service soon:
Documentation of available data not registered and data sources mentioned in other places (e.g. research reports, journal articles).



Data Availability



Further activities to support acquisition

- Increase publicity for archiving activities (e.g. GESIS Road Show).
- Set quality standards for services with Data Seal of Approval.
- Offer hands-on workshops on data management.
- Opening of Secure Data Center for access to sensitive data.

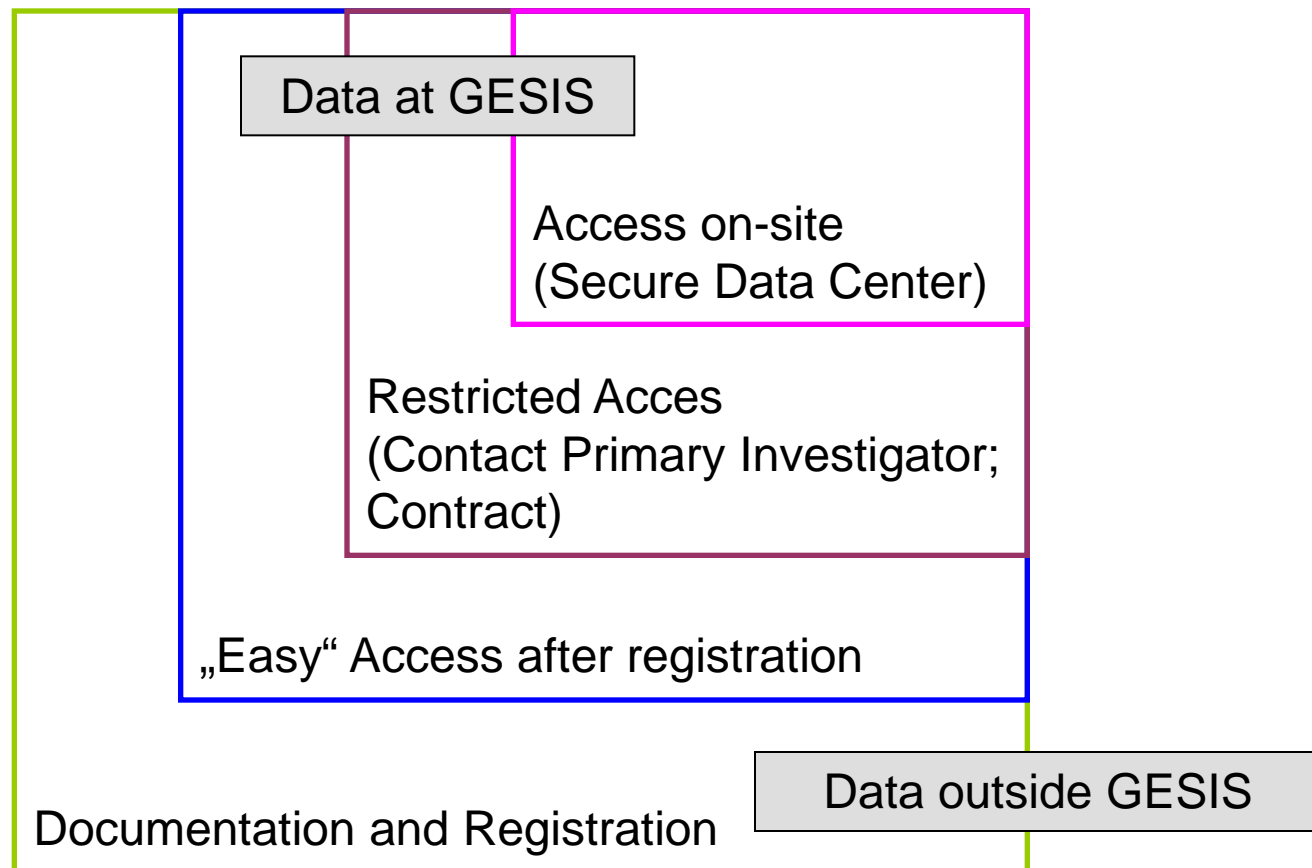
Hands-on Workshops on Data Management

- Team: International Data Infrastructure
- Head: Dr. Alexia Katsanidou
- Preparation for CESSDA ERIC
- Main aim: Make research projects and facilities familiar with the virtues of Data Management Plans.
- Goal: Support preparation of better data.

Secure Data Centre

- Team: Acquisition, Preservation, Data Servicing
- Head: Dr. Katharina Kinder-Kurlanda
- Main aim: Offer a home for sensitive data.
- Goal: Facilitate acquisition of sensitive data kept in the realm of research projects or research facilities for privacy or confidentiality reasons (e.g. elite surveys).

Data Servicing (Access)



A Growing Social Science Data Landscape – A Chance!

- Build and strengthen infrastructure that comprises more than one way to data!
- Facilitate access to data in the (inter)net of resources by archiving AND documentation!
- Unlock new data sources for researchers!

MAP OF THE GOLD REGIONS OF CALIFORNIA. Showing the Routes via Chagres and Panama, Cape Horn, &c.

