SFB882 - INF

Managing and Sharing Data within the Collaborative Research Center SFB882: Concepts and Requirements of a VRE for the Social Sciences

Johanna Vompras

University Library Bielefeld

June 7th, 2012
IASSIST2012 – Washington DC
Session: Infrastructure to Support Research Data Management
Contents

1 Collaborative Research Center SFB882
2 Virtual Research Environment
3 Concepts and functional Requirements
4 Summary
You have never heard about 'Bielefeld'???
Contents

1. Collaborative Research Center SFB882
2. Virtual Research Environment
3. Concepts and functional Requirements
4. Summary
Definitions

- **Heterogeneities**: solely indicate the diversity of members of society. Examination of ascriptive characteristics, like physical differences between people: sex, age, nationality and ethnicity; cultural preferences, attitudes, orientations, and philosophies; differentiation of activities in the scope of division of labor (professions, qualifications, traits).

- **Inequalities**: multidimensional problem: inequality indicates how many individuals have comparable access to different resources (money, prestige, power) and measures how much difference there is between comparable levels of access (e.g. how much effort is required).

Research Questions

How do such heterogeneities lead to social inequalities?  
And what are the social mechanisms that underlie this process?
Research on social Inequalities:
→ Searching for common mechanisms in the emergence of inequality that can be compiled into a typology.

19 sub-projects, in the fields:

<table>
<thead>
<tr>
<th>A</th>
<th>in individual life course (A1 - A6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>in context of organizations (B1 - B5)</td>
</tr>
<tr>
<td>C</td>
<td>beyond the nation-state (C1 - C5)</td>
</tr>
<tr>
<td>Z</td>
<td>central services (Z, MGK, INF)</td>
</tr>
</tbody>
</table>

15 professorships and the University Library,
8,7 millions EUR from 01.07.2011 for a period of four years (initially),
50 senior and junior researchers,
Own integrated Research Training Group (Graduiertenkolleg)
# SFB882: Involved Organizations

## Organizations and Faculties

<table>
<thead>
<tr>
<th>Project Leader</th>
<th>Institute</th>
<th>Sub-Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albert, Mathias, Prof. Dr.</td>
<td>Faculty of Sociology, Bielefeld</td>
<td>C 5</td>
</tr>
<tr>
<td>Dawid, Herbert, Prof. Dr.</td>
<td>Faculty of Economic Sciences, Bielefeld</td>
<td>A 4; MGK</td>
</tr>
<tr>
<td>Diehm, Isabell, Prof. Dr.</td>
<td>Faculty of Education Science, Bielefeld</td>
<td>B 1</td>
</tr>
<tr>
<td>Diewald, Martin, Prof. Dr.</td>
<td>Faculty of Sociology, Bielefeld</td>
<td>A 1; B 3; Z</td>
</tr>
<tr>
<td>Faist, Thomas, Prof. PhD</td>
<td>Faculty of Sociology, Bielefeld</td>
<td>C 1; C 3; Z</td>
</tr>
<tr>
<td>Golsch, Katrin, Dr.</td>
<td>Faculty of Sociology, Bielefeld</td>
<td>A 3</td>
</tr>
<tr>
<td>Gusy, Christoph, Prof. Dr.</td>
<td>Faculty of Law</td>
<td>C 4</td>
</tr>
<tr>
<td>Horstmann, Wolfram, Dr.</td>
<td>University Library, Bielefeld</td>
<td>INF</td>
</tr>
<tr>
<td>Koch, Martin, Dr.</td>
<td>Faculty of Sociology, Bielefeld</td>
<td>C 5</td>
</tr>
<tr>
<td>Liebig, Stefan, Prof. Dr.</td>
<td>Faculty of Sociology, Bielefeld</td>
<td>A 6; B 4; INF</td>
</tr>
<tr>
<td>Lohmann, Henning, Dr.</td>
<td>Faculty of Sociology, Bielefeld</td>
<td>A 5</td>
</tr>
<tr>
<td>Meyer, Christian, Dr.</td>
<td>Faculty of Sociology, Bielefeld</td>
<td>INF</td>
</tr>
<tr>
<td>Müller, Ursula, Prof. Dr.</td>
<td>Faculty of Sociology, Bielefeld</td>
<td>B 5</td>
</tr>
<tr>
<td>Oechsle, Mechtild, Prof. Dr.</td>
<td>Faculty of Sociology, Bielefeld</td>
<td>B 5</td>
</tr>
<tr>
<td>Pfaff-Czarnecka, Joanna, Prof. Dr.</td>
<td>Faculty of Sociology, Bielefeld</td>
<td>B 2</td>
</tr>
<tr>
<td>Razum, Oliver, Prof. Dr. med.</td>
<td>DIW/SOEP, Berlin and Faculty of Soc., Bielefeld</td>
<td>C 1; C 2</td>
</tr>
<tr>
<td>Reinecke, Jost, Prof. Dr.</td>
<td>Faculty of Sociology, Bielefeld</td>
<td>A 2; MGK</td>
</tr>
<tr>
<td>Schupp, Jürgen, Prof. Dr.</td>
<td>Faculty of Health Sciences, Bielefeld</td>
<td>A 1</td>
</tr>
<tr>
<td>Stemmler, Mark, Prof. Ph.D</td>
<td>German Institute for Economic Research (DIW), Berlin</td>
<td>A 2</td>
</tr>
<tr>
<td></td>
<td>Faculty of Psychology, Erlangen-Nürnberg</td>
<td></td>
</tr>
</tbody>
</table>
INF: Information and Data Infrastructure (1)

Motivation: Requirements of the DFG to make primary research data available to research community (10 years).

- Since 2007, it has been possible to apply for funding of service projects, so-called INF projects, within a CRC,
- INF projects aim to develop information infrastructure for the CRCs. Collaboration with infrastructure facilities to combine the expertise of scientists and information specialists:
  - libraries,
  - information centers,
  - media centers, and computing services.
- DFG currently funds 19 INF projects within the framework of more than 200 active CRCs.
INF: Information and Data Infrastructure (2)

1. Work
   - Bundling of IT resources for the subprojects but also for the CRC as a whole,
   - Supporting of (transdisciplinary) research practice,
   - Adapting methods and way of working into VRE.

2. Research Platform
   - Aspects of Research Data Management,
   - Further Development of the socio-scientific methodology,
   - Documentation and Archiving,
   - Re-Use of Research Data and thus replicability of research results.

3. Data Exchange
   - Meta data formats and transformation methods,
   - Interfaces to information resources.
Contents

1 Collaborative Research Center SFB882
2 Virtual Research Environment
3 Concepts and functional Requirements
4 Summary
Definition. VRE.

Virt. Research Environment (VRE)
Concepts and funct. Requirements
Summary

"a platform for internet-based collaborative working that enables new ways of collaboration and a new way of dealing with research data and information ..." [DFG]

"a collaboratory is a web-based ... environment that enables researchers based in different locations to work together and share their knowledge ... speeding up both national and international research." [SURFfoundation]

"[...] managing the increasingly complex range of tasks, [...] shorthand for the tools and technologies [...] The detailed design of a VRE depends of many factors ..." [JISC]
"a platform for internet-based collaborative working that enables new ways of collaboration and a new way of dealing with research data and information [...]" [DFG]
"a platform for internet-based collaborative working that enables new ways of collaboration and a new way of dealing with research data and information [...]”  [DFG]

"a collaboratory is a web-based [...] environment that enables researchers based in different locations to work together and share their knowledge [...] , speeding up both national and international research.”  [SURFfoundation]
"a platform for internet-based collaborative working that enables new ways of collaboration and a new way of dealing with research data and information [...]” [DFG]

"a collaboratory is a web-based [...] environment that enables researchers based in different locations to work together and share their knowledge [...], speeding up both national and international research.” [SURFfoundation]

"[...] managing the increasingly complex range of tasks, [...] shorthand for the tools and technologies [...] ; The detailed design of a VRE depends of many factors …” [JISC]
VRE for SFB882: Aims and Visions
VRE for SFB882: Aims and Visions

Central collection point of projects' output
VRE for SFB882: Aims and Visions

Central collection point of projects’ output

Shorten communication lines between scientists
VRE for SFB882: Aims and Visions

Central collection point of projects’ output

Shorten communication lines between scientists

Upgrade the research work

Data Management

Documentation

Data Publishing
VRE for SFB882: Aims and Visions

- Central collection point of projects’ output
- Shorten communication lines between scientists
- Data Management
- Documentation
- Data Publishing
- Upgrade the research work
- Data Re-Use and Secondary Analysis

Johanna Vompras

Universität Bielefeld
Specific Aspects

- **Data Privacy**: Anonymization techniques (*personal data and data related to company’s business*). No identification of the persons questioned or other individuals/companies mentioned in the interview.

- **Documentation**: Structured, machine-actionable. Documentation throughout the data life cycle. "Quali": interpretations and consideration of primary researchers’ focus.

- **Basic Research**: Methodological research → e.g. Inferring knowledge about the sustainable use of qualitative data which is then used to optimize data collection and analysis.
Generated and used data in the SFB covers almost all types of relevant data for the social sciences.

Types

- Quantitative Microdata
- Qualitative data
  - Interview data (video, audio)
  - Observation data and corresponding documents (inter-linked)
  - Para data (side product of the computer-aided data collection)
- Simulation data (Faculty of Economy)
Contents

1 Collaborative Research Center SFB882
2 Virtual Research Environment
3 Concepts and functional Requirements
4 Summary
Integration and Interoperability

- LDAP-Interface for authentication (user & rights management),
Integration and Interoperability

- LDAP-Interface for authentication (user & rights management),
- Publication management (PUB): institutional repository & academic bibliography,
Integration and Interoperability

- LDAP-Interface for authentication (user & rights mg.),
- Publication management (PUB): institutional repository & academic bibliography,
- Content distribution and linkage of bibliographic data with other resources e.g. BIS (Bielefeld Information System) or PEVZ (Person and Faculty Catalog),
Integration and Interoperability

- LDAP-Interface for authentication (user & rights mng.),
- Publication management (PUB): institutional repository & academic bibliography,
- Content distribution and linkage of bibliographic data with other resources e.g. BIS (Bielefeld Information System) or PEVZ (Person and Faculty Catalog),
- Linkage of research data with publications
Integration and Interoperability

- LDAP-Interface for authentication (user & rights mg.),
- Publication management (PUB): institutional repository & academic bibliography,
- Content distribution and linkage of bibliographic data with other resources e.g. BIS (Bielefeld Information System) or PEVZ (Person and Faculty Catalog),
- Linkage of research data with publications
- Interfaces for third-party producers and distributors of research data, like SOEP and DSZ-BO.
Institution-integrated VRE
Aspects
Evaluation of the requirement analysis
An Arabidopsis thaliana T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics


Authors: Russo, M G.; Li, Y.; Strizhov, N.; Reiss, B.; Dekker, K.; Weiss, D.; Weiss, B.; Meissner, A.

Department: Centrum für Biotechnologie
Fakultät für Biologie
Genomforschung

Abstract: The GABI-Kat population of T-DNA mutagenized Arabidopsis thaliana lines with sequence-characterized insertion sites is used extensively for efficient progress in plant functional genomics. Here we provide details about the establishment of the material, demonstrate the population’s functionality and discuss results from quality control studies. T-DNA insertion mutants of the accession Columbia (Col-0) were created by Agrobacterium tumefaciens-mediated transformation. To allow selection of transformed plants under greenhouse conditions, a sulfadiazine resistance marker was employed. DNA from leaves of T1 plants was extracted and used as a template for PCR-based amplification of DNA fragments spanning insertion site borders. After sequencing, the data were placed in a flanking...
Project 'INFORMIUM' (since April, 2010): Development of a research data infrastructure

Pilot projects started as several collaborations between scientists and infrastructure facilities with the aim to determine **functional requirements** for the realization of a **transdisciplinary research data infrastructure**.

- Advisory services for research data management (e.g. DMP, basic and legal issues, knowledge-transfer)
- Reliable archiving, publication and distribution of research data
Requirements Analysis

VRE highly dependent on requirements of the researchers working within the SFB

- Analysis and determination of different:
  - Working and collaboration workflows,
  - Research methods,
  - Types of research data within the subprojects and their requirements regarding
    - documentation
    - data security
    - processing, archiving and re-use

Survey

- Practices are defined among researchers of one sub-project,
- Participation of one person per sub-project.
Aspects:

1. Work collaboration

How do you communicate with your colleagues?
How do you collaboratively create a bibliography?

Which text-processing software do you use?
How do you import bibliography items into your database?

Where do you store the corresponding publications?
Do you process sensitive data in your project?

Workflows for publishing data from VRE on external SFB website
Aspects:

1. **Work collaboration**
   - How do you communicate with your colleagues?
Aspects:

1. **Work collaboration**
   - How do you communicate with your colleagues?
   - How do you collaboratively create a bibliography?
Aspects:

1. Work collaboration
   - How do you communicate with your colleagues?
   - How do you collaboratively create a bibliography?

2. Tools/Software
Aspects:

1. Work collaboration
   - How do you communicate with your colleagues?
   - How do you collaboratively create a bibliography?

2. Tools/Software
   - Which text-processing software do you use?
Aspects:

1. Work collaboration
   - How do you communicate with your colleagues?
   - How do you collaboratively create a bibliography?

2. Tools/Software
   - Which text-processing software do you use?
   - How do you import bibliography items into your database?
Aspects:

1. Work collaboration
   - How do you communicate with your colleagues?
   - How do you collaboratively create a bibliography?

2. Tools/Software
   - Which text-processing software do you use?
   - How do you import bibliography items into your database?

3. Data archiving
Aspects:

1. Work collaboration
   - How do you communicate with your colleagues?
   - How do you collaboratively create a bibliography?

2. Tools/Software
   - Which text-processing software do you use?
   - How do you import bibliography items into your database?

3. Data archiving
   - Where do you store the corresponding publications?
Aspects:

1. Work collaboration
   - How do you communicate with your colleagues?
   - How do you collaboratively create a bibliography?

2. Tools/Software
   - Which text-processing software do you use?
   - How do you import bibliography items into your database?

3. Data archiving
   - Where do you store the corresponding publications?
   - Do you process sensitive data in your project?
Aspects:

1. Work collaboration
   - How do you communicate with your colleagues?
   - How do you collaboratively create a bibliography?

2. Tools/Software
   - Which text-processing software do you use?
   - How do you import bibliography items into your database?

3. Data archiving
   - Where do you store the corresponding publications?
   - Do you process sensitive data in your project?

4. Global institutional services
Aspects:

1. Work collaboration
   - How do you communicate with your colleagues?
   - How do you collaboratively create a bibliography?

2. Tools/Software
   - Which text-processing software do you use?
   - How do you import bibliography items into your database?

3. Data archiving
   - Where do you store the corresponding publications?
   - Do you process sensitive data in your project?

4. Global institutional services
   - Workflows for publishing data from VRE on external SFB website
Results: Collaboration

- Various types of files,
- Directory structure in the VRE, linking to institutional hard drives,
- Notifications for composing and changing,
- Collaborative collection of data (e.g., publications, notes),
- Conceptualization of work (keywords, cv, free annotation),
- Real-time messaging,
- Calendar (private, group), mobile devices.
Results: Tools, e.g Bibliography Management

- User and rights management,
- Managing own and project-related publications,
- Publishing and linking,
- Import and export functions,
- Import by Persistent Identifier, PUB-ID, BibTeX, EndNote, etc.
- Various citation styles.
Results: Data Manipulation

- Textprocessing (real-time collaborating),
- Automat. capture Metadata,
- DDI3 for parts of the life-cycle,
- Document/Notes/Data linkage,
- Format-specific transformations, conversions,
- Processing statistical calculations,
- Versioning, 'track changes'.

Johanna Vompras
IAASSIST2012, June 4th – 8th, Washington D.C.
Results: Data Archiving

- Research Data Management,
- Re-Use,
- Data Encryption,
- Redundant Storage, Backup.
Evaluation of the requirement analysis

Institution-integrated VRE Aspects

Johanna Vompras
IASSIST2012, June 4th – 8th, Washington D.C.
Contents

1 Collaborative Research Center SFB882
2 Virtual Research Environment
3 Concepts and functional Requirements
4 Summary
Summary and Conclusions

- VRE is highly dependent on the researchers’ needs and work processes,
- Support the processes of conducting research, but also ensure high-quality data documentation and long-term archiving of the project output,
- VRE for SFB882 is work in progress,
- Current status: definition and implementation of core components for VRE, like
  - Collaboration: WIKI, Calendar, Notifications,
  - Data Storage: File Upload, File Linkage,
  - Integration of VRE to institutional services.
- Future Work: DDI-based Data Documentation, Step-by-Step adjusting of ”work-spaces” to the subprojects’ needs.
Tools: R-Server
Tools: Task Manager
Tools: Desktop

---

Summary

University Bielefeld

SFB882

Virt. Research Environment (VRE)

Concepts and funct. Requirements

Summary

Optional: Screenshots

---

Johanna Vompras

IAASSIST2012, June 4th – 8th, Washington D.C.
References

- University Library Bielefeld
  http://ub.uni-bielefeld.de

- SFB882 INF
  http://sfb882.uni-bielefeld.de/projects/Teilprojekt_INF

- Data Service Center for Business and Organizational Data (DSZBO):
  http://www.uni-bielefeld.de/soz/datenservicezentrum/
The End.

Thank you for your attention!

Dr. Johanna Vompras
University Library Bielefeld
Research Data Management Services

johanna.vompras@uni-bielefeld.de