

## *Impulsvortrag*

Datenmanagement und Bereitstellung von Werkzeugen - technische Herausforderungen der INF-Projekte?

am Beispiel des SFB882

Johanna Vompras

Universität Bielefeld

11. April 2013

Treffen der INF-Projekte – SUB Göttingen

# Contents

- 1 Sonderforschungsbereich SFB882
- 2 Der sozialwissenschaftliche Forschungsprozess
- 3 INF
- 4 Technische Herausforderungen

# Inhaltsverzeichnis

- 1 Sonderforschungsbereich SFB882
- 2 Der sozialwissenschaftliche Forschungsprozess
- 3 INF
- 4 Technische Herausforderungen

# SFB882: Von Heterogenitäten zu Ungleichheiten

## Allgemeines

- **Heterogenitäten:** indizieren zunächst eine bloße Verschiedenheit von Gesellschaftsmitgliedern. Untersuchung von askriptiven Merkmalen wie physischen Unterschieden zwischen den Menschen, Geschlecht, Alter, Nationalität und Ethnizität; kulturelle Vorlieben oder Lebensformen, sowie Differenzierung von Tätigkeiten im Rahmen der gesellschaftlichen Arbeitsteilung.
- **Ungleichheiten:** mehrdimensionales Phänomen. Perspektiven: z.B.: Wohlbefinden, Beteiligungschancen in unterschiedlichen Lebensbereichen und der nötiger Anstrengung.

## Erkenntnisinteresse

Untersuchung sozialer Mechanismen, durch die individuelle Heterogenitäten im Sinne bloßer Verschiedenheit der Gesellschaftsmitglieder zu sozialen Ungleichheiten führen.

## SFB882: Beteiligte Organisationen

## Organisationen und Fakultäten

**Project Leader**

Albert, Mathias, Prof. Dr.  
 Dawid, Herbert, Prof. Dr.  
 Diehm, Isabell, Prof. Dr.  
 Diewald, Martin, Prof. Dr.  
 Faist, Thomas, Prof. PhD  
 Golsch, Katrin, Dr.  
 Gusy, Christoph, Prof. Dr.  
 Horstmann, Wolfram, Dr.  
 Koch, Martin, Dr.  
 Liebig, Stefan, Prof. Dr.  
 Lohmann, Henning, Dr.  
 Meyer, Christian, Dr.  
 Müller, Ursula, Prof. Dr.  
 Oechsle, Mechtild, Prof. Dr.  
 Pfaff-Czarnecka, Joanna, Prof. Dr.  
 Razum, Oliver, Prof. Dr. med.  
 Reinecke, Jost, Prof. Dr.  
 Schupp, Jürgen, Prof. Dr.  
 Stemmler, Mark, Prof. Ph.D

**Institute**

Faculty of Sociology, Bielefeld  
 Faculty of Economic Sciences, Bielefeld  
 Faculty of Education Science, Bielefeld  
 Faculty of Sociology, Bielefeld  
 Faculty of Sociology, Bielefeld  
 Faculty of Sociology, Bielefeld  
 Faculty of Law  
 University Library, Bielefeld  
 Faculty of Sociology, Bielefeld  
 Faculty of Sociology, Bielefeld  
 DIW/SOEP, Berlin and Faculty of Soc., Bielefeld  
 Faculty of Sociology, Bielefeld  
 Faculty of Sociology, Bielefeld  
 Faculty of Sociology, Bielefeld  
 Faculty of Sociology, Bielefeld  
 Faculty of Health Sciences, Bielefeld  
 Faculty of Sociology, Bielefeld  
 German Institute for Economic Research (DIW), Berlin  
 Faculty of Psychology, Erlangen-Nürnberg

**Sub-Project**

C 5  
 A 4; MGK  
 B 1  
 A 1; B 3; Z  
 C 1; C 3; Z  
 A 3  
 C4  
**INF**  
 C5  
 A 6; B 4; INF  
 A 5  
 INF  
 B 5  
 B 5  
 B 2  
 C 1, C 2  
 A 2; MGK  
 A 1  
 A 2

# Inhaltsverzeichnis

- 1 Sonderforschungsbereich SFB882
- 2 Der sozialwissenschaftliche Forschungsprozess
- 3 INF
- 4 Technische Herausforderungen

# Forschungsdaten in den Sozialwissenschaften (1)

*'Valid and reliable data are indispensable foundation for research in the social sciences and economics: they ensure that research is in line with the contemporary realities and provide convincing arguments for actions by citizens, policy-makers, and business leaders'<sup>1</sup>*

---

<sup>1</sup>Auszug aus einem Report der KVI, Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik, 2001. 

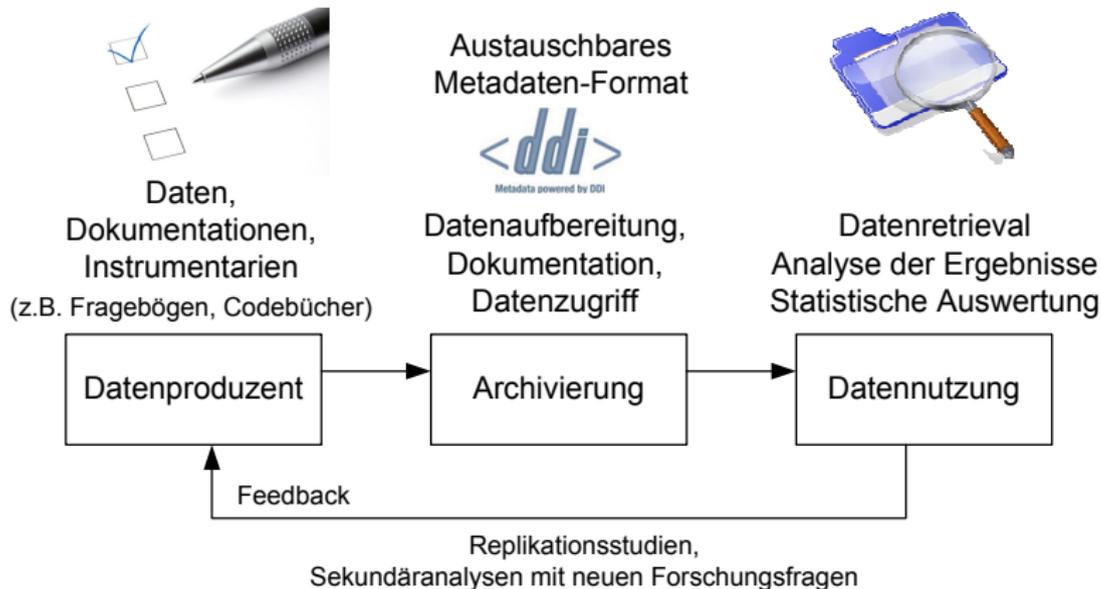
# Forschungsdaten in den Sozialwissenschaften (2)

## Mehrere Ebenen

z.B. quantitative Forschung: Verhalten wird durch Modelle, Zusammenhänge und zahlenmäßige Ausprägungen möglichst genau beschrieben und vorhersagbar gemacht

Studie Erhebung	Allgemeine Beschreibung Methodenbericht z.B. wiss. Fragestellung, Datenerhebungsmodus Fragebogenerstellung, Feldphase, Datenweitergabe etc.
Fragebogenfragen Variablen Ausprägungen	Fragebogen, Codebuch Codebuch: Variablennamen, Variablenlabels Codebuch: Ausprägungsnamen (1,2,3,...), Ausprägungslabls, Fehlende Werte
Datensätze	Daten selbst (Mikrodaten)

## ... und ihre Verarbeitung



# Inhaltsverzeichnis

- 1 Sonderforschungsbereich SFB882
- 2 Der sozialwissenschaftliche Forschungsprozess
- 3 INF**
- 4 Technische Herausforderungen

# INF: Aufbau der Informations- und Dateninfrastruktur

## 1 Arbeitsplattform

- Bündelung von IT-Ressourcen für Teilprojekte und Gesamt-SFB
- Unterstützung bei der Forschungspraxis
- Adaptation an die Arbeitsweisen

## 2 Forschungsplattform

- Aspekte des Forschungsdatenmanagements
- Dokumentation und Datenarchivierung
- Nachnutzung von Forschungsdaten

## 3 Datenaustausch

- Spezielle Metadatenformate
- Schnittstellen zu existierenden Informationsressourcen
- Transformationsfunktionalitäten

# Data Documentation Initiative (DDI)<sup>2</sup>

- XML-basierter Metadaten-Standard zur Dokumentation,
- *Version 3.0 seit 2004, Version 3.1 seit 2009, 3.2 ab 2012,*
- Datendokumentation über den gesamten *Data Life Cycle*,
- Datenobjekte werden als Module und Schemas beschrieben:
  - Re-Use und Versionierung durch Identifier
  - Mehrsprachlichkeit

## Hauptziel

- Dokumentation → Wiederverwendbarkeit (Re-Use)
- Austausch (Data Exchange)
- Erschließung von Zusammenhängen / Vergleichbarkeit von Studien

<sup>2</sup><http://www.ddialliance.org>

# Datenlebenszyklus (*data life cycle model*)

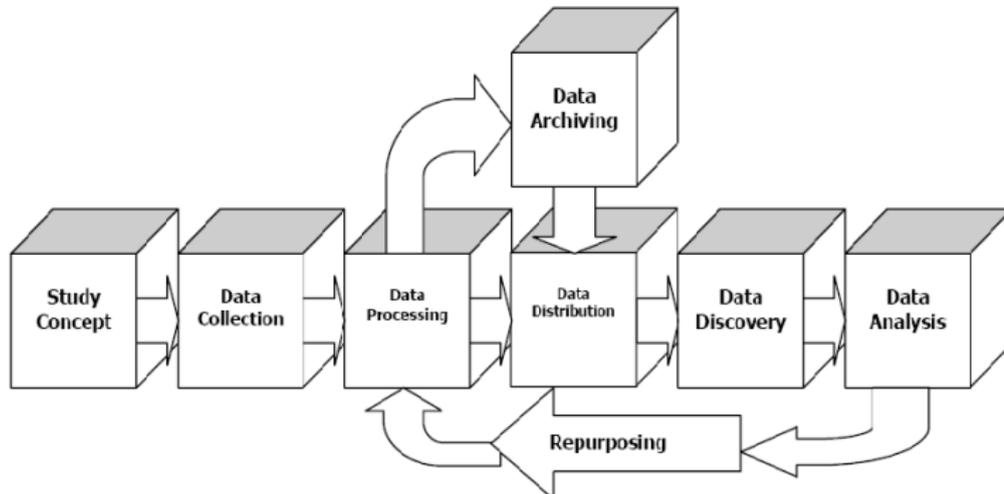


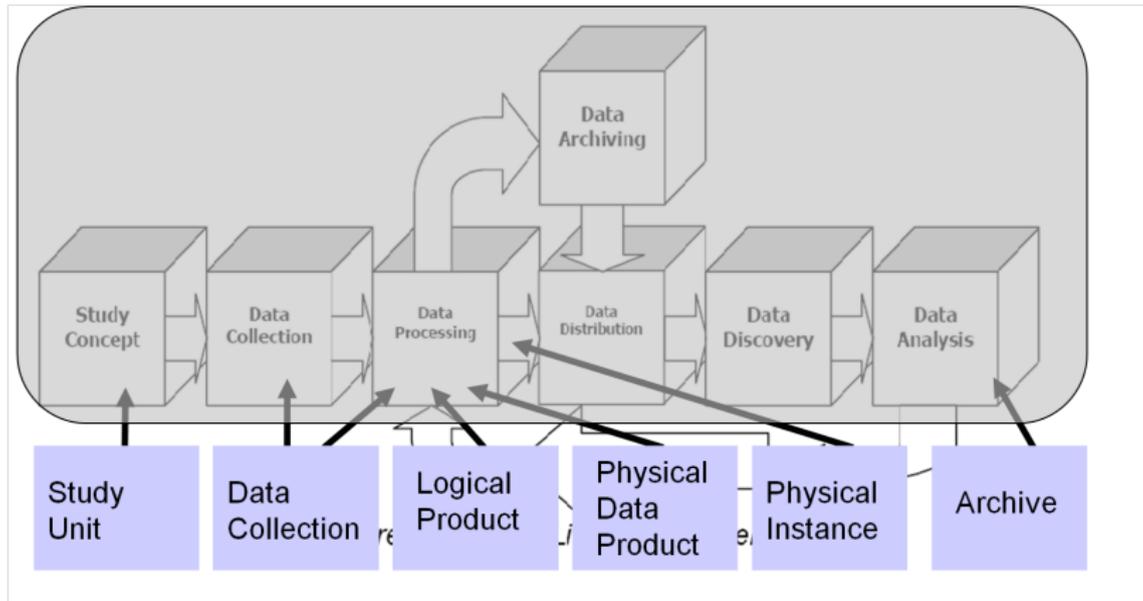
Figure: Combined Life Cycle Model

Quelle: DDI Alliance. *DDI Version 3 Conceptual Model (Draft 2004)*.

# Beispiel einer DataCollection in DDI3.x

```
<d:DataCollection id="DataCollection_4245" version="4.0">
  <d:QuestionScheme id="QuestionScheme_4245" version="2.0">
    <d:QuestionItem id="A7">
      <d:QuestionText xml:lang="en">
        <d:LiteralText>
          <d:Text>How many days in the PAST WEEK did you watch NATIONAL network news on TV?
        </d:Text>
        </d:LiteralText>
      </d:QuestionText>
      <d:CodeDomain>
        <r:CodeSchemeReference>
          <r:ID>CodeScheme_V043014</r:ID>
        </r:CodeSchemeReference>
      </d:CodeDomain>
    </d:QuestionItem>
    ...
  </d:QuestionScheme>
  ...
</d:DataCollection>
```

# Mapping von StudyUnit in DDI



# Inhaltsverzeichnis

- 1 Sonderforschungsbereich SFB882
- 2 Der sozialwissenschaftliche Forschungsprozess
- 3 INF
- 4 Technische Herausforderungen**

# Tools



# Bedarfsanalyse

Ermittlung funktionaler Anforderungen für die technische Umsetzung der Arbeitsumgebung:

- Hochgradig abhängig von den Anforderungen der Wissenschaftler im SFB,
- Analyse und Ermittlung von unterschiedlichen
  - Arbeitsweisen,
  - Forschungsmethoden,
  - Arten von Forschungsdaten innerhalb der Teilprojekte,

## Vorgehen

- Eine Person pro Projekt nimmt teil,
- Arbeitsweise wird je Projekt definiert.

# Technische Konzepte zur Maximierung der Akzeptanz

- Anpassung der Arbeitsumgebung an die gewohnten Arbeitsweisen der Forschenden, z.B. Einbindung von Shared Laufwerken
- Nutzung intuitiver Tools, z.B. Dateiupload, Versionierung, "One-Click"-Publikation von Forschungsdaten
- Integration von bisher genutzten Workflows (z.B. Anmeldung LDAP, Publikationsdienste)
- Ständiges Abwägen zwischen Nutzen/Mehraufwand für den einzelnen Forschenden
- Den "Mehrwert" automatisiert während der Arbeitsprozesse extrahieren (z.B. Verknüpfung Zeit/Daten; Versionen, Dokumentation direkt beim Upload anbieten, etc.)

# Interoperabilität

Sicherstellung von Interoperabilität zwischen der VRE und globalen universitären Diensten:

- LDAP-Anbindung zur Authentifizierung (User & Rechtemanagement)
- Nutzung von der im PEVZ gespeicherten Personendaten
- Einbettung Publikationsmanagement (PUB System)
- Webauftritt SFB882 / Projektinformationen / Personen
- Schnittstellen zu Drittanbietern von Forschungsdaten
  - Ein- und Ausgabefunktionen zum Datenaustausch
  - SOEP und andere Archive, wie FDZ und DSZ.

# Herausforderungen bei der Toolentwicklung

- Übergreifender Überblick über den Bedarf und Anforderungen

# Herausforderungen bei der Toolentwicklung

- Übergreifender Überblick über den Bedarf und Anforderungen  
→ Rückkopplung mit Projekten

# Herausforderungen bei der Toolentwicklung

- Übergreifender Überblick über den Bedarf und Anforderungen  
→ Rückkopplung mit Projekten
- Heterogenitäten der Daten und Forschungsprozesse

# Herausforderungen bei der Toolentwicklung

- Übergreifender Überblick über den Bedarf und Anforderungen  
→ Rückkopplung mit Projekten
- Heterogenitäten der Daten und Forschungsprozesse  
→ Anpassung der Tools zur Dokumentation und Workflows zur nachhaltigen Archivierung

# Herausforderungen bei der Toolentwicklung

- Übergreifender Überblick über den Bedarf und Anforderungen  
→ Rückkopplung mit Projekten
- Heterogenitäten der Daten und Forschungsprozesse  
→ Anpassung der Tools zur Dokumentation und Workflows zur nachhaltigen Archivierung
- Überwachung der in den Fachdisziplinen gängigen Standards und Best Practices

# Herausforderungen bei der Toolentwicklung

- Übergreifender Überblick über den Bedarf und Anforderungen  
→ Rückkopplung mit Projekten
- Heterogenitäten der Daten und Forschungsprozesse  
→ Anpassung der Tools zur Dokumentation und Workflows zur nachhaltigen Archivierung
- Überwachung der in den Fachdisziplinen gängigen Standards und Best Practices  
→ z.B. Weiterentwicklung der Dokumentationsstandards

# Herausforderungen bei der Toolentwicklung

- Übergreifender Überblick über den Bedarf und Anforderungen  
→ Rückkopplung mit Projekten
- Heterogenitäten der Daten und Forschungsprozesse  
→ Anpassung der Tools zur Dokumentation und Workflows zur nachhaltigen Archivierung
- Überwachung der in den Fachdisziplinen gängigen Standards und Best Practices  
→ z.B. Weiterentwicklung der Dokumentationsstandards
- Minimierung des *"Akzeptanzproblems"*

# Herausforderungen bei der Toolentwicklung

- Übergreifender Überblick über den Bedarf und Anforderungen  
→ Rückkopplung mit Projekten
- Heterogenitäten der Daten und Forschungsprozesse  
→ Anpassung der Tools zur Dokumentation und Workflows zur nachhaltigen Archivierung
- Überwachung der in den Fachdisziplinen gängigen Standards und Best Practices  
→ z.B. Weiterentwicklung der Dokumentationsstandards
- Minimierung des *"Akzeptanzproblems"*  
→ d.h. Tools, die nicht belasten (Ressourcen), sondern unterstützen

# Herausforderungen bei der Toolentwicklung

- Übergreifender Überblick über den Bedarf und Anforderungen  
→ Rückkopplung mit Projekten
- Heterogenitäten der Daten und Forschungsprozesse  
→ Anpassung der Tools zur Dokumentation und Workflows zur nachhaltigen Archivierung
- Überwachung der in den Fachdisziplinen gängigen Standards und Best Practices  
→ z.B. Weiterentwicklung der Dokumentationsstandards
- Minimierung des *"Akzeptanzproblems"*  
→ d.h. Tools, die nicht belasten (Ressourcen), sondern unterstützen
- Konzeptualisierung und Implementierung stark ausgerichtet an Übergabe an Grundausstattung

# Herausforderungen bei der Toolentwicklung

- Übergreifender Überblick über den Bedarf und Anforderungen  
→ Rückkopplung mit Projekten
- Heterogenitäten der Daten und Forschungsprozesse  
→ Anpassung der Tools zur Dokumentation und Workflows zur nachhaltigen Archivierung
- Überwachung der in den Fachdisziplinen gängigen Standards und Best Practices  
→ z.B. Weiterentwicklung der Dokumentationsstandards
- Minimierung des *"Akzeptanzproblems"*  
→ d.h. Tools, die nicht belasten (Ressourcen), sondern unterstützen
- Konzeptualisierung und Implementierung stark ausgerichtet an Übergabe an Grundausstattung  
→ Einbeziehung HRZ/UB/FFT, IT Strategie

# Herausforderungen bei der Toolentwicklung

- **Übergreifender Überblick über den Bedarf und Anforderungen**  
→ Rückkopplung mit Projekten
- **Heterogenitäten der Daten und Forschungsprozesse**  
→ Anpassung der Tools zur Dokumentation und Workflows zur nachhaltigen Archivierung
- **Überwachung der in den Fachdisziplinen gängigen Standards und Best Practices**  
→ z.B. Weiterentwicklung der Dokumentationsstandards
- **Minimierung des "Akzeptanzproblems"**  
→ d.h. Tools, die nicht belasten (Ressourcen), sondern unterstützen
- **Konzeptualisierung und Implementierung stark ausgerichtet an Übergabe an Grundausstattung**  
→ Einbeziehung HRZ/UB/FFT, IT Strategie  
→ Einbeziehung EDV Fakultät

# Referenzen

- SFB882  
<http://sfb882.uni-bielefeld.de>
- SFB882 INF  
<http://sfb882.uni-bielefeld.de/projects/inf>
- Forschungsdatenmanagement an der Universität Bielefeld  
<http://data.uni-bielefeld.de>

Ende.

Danke für Ihre Aufmerksamkeit!

Dr. Johanna Vompras

University Library Bielefeld

Research Data Management Services

[johanna.vompras@uni-bielefeld.de](mailto:johanna.vompras@uni-bielefeld.de)