



# ■ Forschungsdatenmanagement in disziplinären Verbundprojekten – eine Praxisreflexion

**Peter Bartelheimer**

**5. Workshop des AK Forschungsdaten der  
Leibniz-Gemeinschaft  
Berlin, 28. Juni 2016**

# ■ Projekthintergrund – auch interdisziplinär

## ■ **Berichterstattung zur sozioökonomischen Entwicklung in Deutschland**

- Drei vom BMBF geförderte Forschungsverbundvorhaben
- Derzeit (2013 bis 2016) : Dritter Bericht (*soeb 3*): vier Universitäten, drei Forschungsdatenzentren, zehn weitere Einrichtungen / Personen

## ■ **Virtuelle Forschungsumgebung für die Sozioökonomische Berichterstattung (VFU *soeb 3*)**

- BMBF-Verbundprojekt 2012 bis 2014: neun beteiligte Einrichtungen

## ■ **Verbundvorhaben der Evaluationsforschung, z.B.**

- Evaluation der Leistungen für Bildung und Teilhabe im Auftrag des BMAS 2013 bis 2016: drei Teilprojekte, im Teilprojekt Implementation sechs Forschungseinrichtungen / Personen
- Evaluation der Beratungskonzeption der Bundesagentur für Arbeit 2011 bis 2014: fünf Forschungseinrichtungen / Personen

# ■ Forschungsdatenmanagement als unterschätzte Aufgabe

## ■ Projektmanagement, Forschungsverbundmanagement hat kaum fachliches Profil

- Bisher kein Standard für Forschungsdatenmanagement, Projektleitung oder studentische Hilfskraft macht es »nebenher«
- Wachsende Bedeutung bei Projektanträgen, -angeboten – aber wie viel darf es kosten?
- An Forschungseinrichtungen fehlt häufig ein Fachkonzept, Aufgabe erfährt keine Aufmerksamkeit auf Leitungsebene

## ■ Learning by Doing: Wo lernt man Forschungsdatenmanagement

## ■ Vorsicht »Hundekurve« – Verbundprojekte können zu Beginn viel Zeit mit Aushandlung verlieren

- Nicht verabreden, vorgeben. Aber wo kommen gute Vorgaben her?
- Entlastend: IT-Unterstützung für Managementfunktionen einbringen

# ■ Forschungsdaten (1): Faktisch oder schwach anonymisierte quantitative Mikrodaten

## ■ In der Regel Nachnutzung existierender Studien

- Bevölkerungssurveys (Querschnitt- oder Paneldesign)  
Prozessdaten der Sozialleistungsträger

## ■ Kooperationsbedingungen

- Daten liegen lokal verteilt
- Nur ein Teil der Arbeitspakete arbeitet mit gleichen Studien
- Z. Tl. unterschiedliche Versionen der gleichen Studie
  - SUF, schwächer anonymisierte Daten in FDZ
- Datenvolumen nur ausnahmsweise ein Problem
- Kosten der Datenbeschaffung – in der Regel auskömmlich kalkuliert

## ■ Vertragsfragen

- Unterschiedlich gestaltete Nutzungsverträge und Berechtigungen

# ■ Forschungsdaten (2): Qualitative Daten

## ■ In der Regel im Rahmen des Projekts erhoben

- Interviews, Beobachtungsprotokolle, Gesprächsbeobachtungen, Falldarstellungen

## ■ Z. Tl. übermittelte personenbezogene Daten

- Fallakten (digital und analog)

## ■ Vertragsfragen

- Datenschutzrechtliche Zusatzvereinbarungen, Anträge auf Übermittlung von Sozialdaten nach § 75 SGB X, Datenübermittlungsverträge
- Vertragsgestaltung je Projekt verschieden
- Oft faktischer Kontrahierungszwang,
- Unklare Begriffe, z.B. Sozialdaten, Übermittlung

# ■ Forschungsdaten (3): Löschpflichten und Archivierung

## ■ Datenarchivierung beim Datengeber, Auftraggeber

- Spannungsreiche Kopplung mit Löschfristen
- Selbst erhobene qualitative Daten: Unklares Verhältnis von (Sozial-) Datenschutz, Urheberrecht, Nutzungsrecht; Kontroverse Bedingungen für Nachnutzung

## ■ Lokale Archivierung

- Keine Standards für (Langzeit-) Archivierung, keine Nachnutzung
- DOI spielen bisher kaum eine Rolle
- Richtungweisendes Projekt eLabour
  - BMBF-geförderter Digital-Humanities-Verbund in Göttingen:
  - Aufbau eines Zentrums für IT-basierte Sekundäranalyse arbeitssoziologischen Materials

# ■ Kommunikation und Datenaustausch bisher

## ■ Kommunikation im Projekt

- Arbeitstreffen: »Königsdisziplin«, teuer und selten professionell
- Synchron und nie stabil: Telefon- und Videokonferenzen
- Email als Standard
- Fortgeschritten: Wiki, Forum,

## ■ IT- unterstützter Datenaustausch

- Freigegebene Dropbox- oder Cloud-Verzeichnisse
- Eigene ftp-Server, Datenaustauschplattform der BA
- BSCW-Server der Bundesanstalt für IT-Dienstleistungen
- Zentrale Datenablage mit Versionsverwaltung, Annotation, Benachrichtigung, Bearbeitung

## ■ **Gemeinsamer »Workflow« bisher**

### ■ **Individueller Stil setzt sich meist durch bei**

- Bei zentraler Datenablage schwierige Verständigung auf Ablagelogik
- Versionierung, Namenskonventionen, Dateiorganisation: »Compliance«, Durchsetzung bleibt schwierig

### ■ **Bei Analyse quantitativer Daten sind gemeinsame Workflows Ausnahme**

- Kommunikation über Ergebnisse
- Individueller Stil der Dokumentation von Syntax
- Nutzung von Dateneinrichtungen erzwingt Mindeststandards

### ■ **Kollaboration bei Analyse qualitativer Daten**

- Auswertungstechniken: Codierung, Memos, Inventare
- Codierung und Auswertung



# ■ Virtuelle Forschungsumgebung (VFU) als neue Qualität

## ■ Virtuelle Forschungsumgebung für die Sozial- und Wirtschaftswissenschaften (VFU SuW)

- Operativer Prototyp für Sozioökonomische Berichterstattung (VFU soeb 3)
- Demoversion für Nachnutzung: <http://vfu-demo.sofi.gwdg.de>

## ■ Funktionsbereiche

- Gemeinsame Oberfläche mit Authentifizierung und rollenbasierter Autorisierung
- Projekt-, Verbundmanagement: Informationen, Verabredungen, Kalender
- Kollaborative Tools: Forum, Wiki, Dateiverwaltung
- Forschungsdatentools: Editoren zur Archivierung, Dokumentation und Nachnutzung von Forschungsdaten und Auswertungssyntax

# ■ Erste Erfahrungen mit Datenmanagement in der VFU

## ■ Kollaborative Funktionen werden intensiv genutzt

- Wiki stärker angenommen als Forum, Chat spielt keine Rolle
- Suchfunktionen entscheidend für Akzeptanz
- Offene Frage: Welche Metadaten und wie viele

## ■ Syntax Sharing bleibt die Ausnahme

- Arbeitsprozesse, Arbeitsstände bleiben privat

## ■ IT setzt Standards durch Verhaltensanforderungen

- Arbeitsweisen reagieren auf technische Möglichkeiten, sind nicht technisch definiert
- Kommunikation und Nutzung bleibt Aushandlungssache und Lernprozess

# ■ Überlegungen zur Weiterentwicklung: Unterstützung von »Remote Data Access«

## ■ Forschungsdaten einbinden – aber wie? Ein Lernprozess.

- Angenommener Nutzungsfall ursprünglich:
  - Ablage von Forschungsdaten in der VFU, Integration von Statistikprogrammen (lizenzfrei: R)
- Annahmen heute:
  - Stark anonymisierte Forschungsdaten in VFU, gerechnet wird lokal
  - Schwächer anonymisierte Daten verbleiben in FDZ, gerechnet wird dort

## ■ Technische Entwicklungsrichtungen

- Unterstützung von Verfahren des Fernrechnens in FDZ
- Einbindung eines »Remote Desktop«, »Remote Data Access«

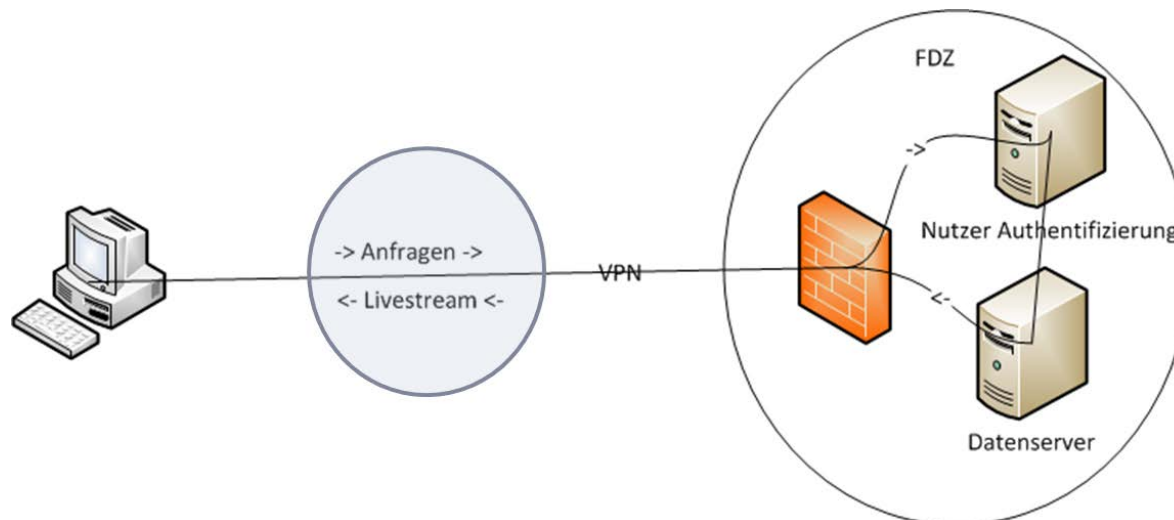
## ■ Backup

# ■ Abgesicherter Datenfernzugriff – Proof of concept

## ■ Nutzung der VFU als Verfahren der Zugangsbeschränkung

- («single point of access«)
- Ggf. Nutzung an gesichertem Zugangsort («safe room«)
- Beschränkung des Datenzugangs auf vertraglich erlaubte Studien

## ■ Integration eines sicheren »VFU-Arbeitsplatzes« auf externen Servern



# ■ Leistungen eines »Remote Desktop« in FDZ

- Zugriff auf Daten, für die Berechtigung besteht
- Nutzung von Statistikprogrammen (R, stata, SPSS usw.)
- Verzeichnisse für Ablage am virtuellen Arbeitsplatz
  - Z.B. Ergebnisdateien, Syntax, Logs
- Browsen und Bearbeitung in Echtzeit (Livestream)
  - Sicherer Transfer von Syntax und Output
  - Output-Kontrolle beschränkt auf zu verwendende Ergebnisse
- Anwendungsfälle
  - »Remote Desktop« als Alternative zu Gastarbeitsplätzen in FDZ und gesicherten Arbeitsplätzen in Forschungseinrichtungen
  - VFU als Arbeitsumgebung für Nutzung der Dateninfrastruktur