

■ **Peter Bartelheimer, Tanja Schmidt**
Diesseits der Datenschnittstelle –
Nutzung von Forschungsdaten
in virtuellen Arbeitsumgebungen

5. Konferenz für Sozial- und Wirtschaftsdaten (KSWD)
Wiesbaden, 13./14. Januar 2011

■ Integrierte digitale Forschungsumgebungen – Innovation über die ganze Leistungskette

■ Leitbild der Allianz-Initiative Digitale Information (2008)

- Lizenzierung / Hosting: publiziertes Wissen nachnutzbar machen
- Sicherung, Archivierung und Bereitstellung von Forschungsprimärdaten
- *Virtuelle Forschungsumgebungen (Aktionsfeld 5)*
- Rechtlicher Rahmen: z.B. Urheberrechte, Datenschutz

■ Kommission Zukunft der Informationsinfrastruktur (KII)

- Arbeitsgruppe Virtuelle Forschungsumgebungen

■ Was eine virtuelle Forschungsumgebung leisten soll

- »Der Forschung alle nötigen Instrumente, Daten, Informationen und Werkzeuge ohne Ressourcen- und Zugangsprobleme zur Verfügung stellen« (Neuroth / Aschenbrenner / Lohmeier 2007)

■ Future Data Access *und* virtuelle Forschungsumgebungen

- Innovative Informationstechnik beiderseits der Datenschnittstelle

■ Virtuelle Forschungsumgebungen – was bisher geschah?

■ BMBF: eScience im Förderprogram IKT 2020

- D-Grid-Initiative: Infrastruktur für Grid-Computing
- Projekt WissGrid (Koordination Ernst-August-Universität Göttingen): Unterstützung neuer wissenschaftlicher Disziplinen bei Nutzung standortübergreifender IT-Ressourcen
 - Gemeinsame Vertretung gegenüber Gauß-Allianz (D-Grid-Support)
 - Blaupausen und Fachberater für wissenschaftliche Communities
 - Langzeitarchivierung digitaler Forschungsdaten

■ DFG: Themenschwerpunkt Digitale Information

- Förderbereich Informationsmanagement
- Aktionslinie 13: Virtuelle Forschungsumgebungen
- Ausschreibung Informationsinfrastrukturen für Forschungsdaten (2010)

- **Modellprojekt Kollaborative Datenauswertung und virtuelle Arbeitsumgebung**

- **Entwicklung einer virtuellen Forschungsumgebung für quantitative Sozialforschung am »Anwendungsfall«**
 - Zweiter Forschungsverbund »Berichterstattung zur sozioökonomischen Entwicklung Deutschlands« 2005 – 2009, Bericht im Erscheinen
 - In 2010 fachöffentliche Werkstattgespräche zu einem dritten Bericht
 - Koordination: SOFI Göttingen

- **Modellprojekt Virtuelle Arbeitsumgebung – vier Workshops 2010**
 - Workshop 1: Fachwissenschaftliche Anforderungen
 - Workshop 2: Evaluation der Grid-Technologie, Architekturskizze
 - Workshop 3: Rechtliche Aspekte der Nutzung von Forschungsdaten
 - Workshop 4: Diskussion von Projektergebnissen mit Anwender/inne/n

- **Kooperation:**
 - Wiss-Grid-Fachberater, D-Grid-GmbH, Forschungsdatenzentren

■ Hauptfunktionen in einer virtuellen Forschungsumgebung (1)

■ Datenverwaltung

- Forschungsdatensätze, Syntaxdateien und Metadaten
- Suchfunktionen für alle Datenarten, Datenversionierung, Verknüpfungen
- Individuelle Nutzungsrechte und –Zeiträume durch Zutritt zu virtuellen Organisationen

■ Datenbearbeitung

- Editieren von Metadaten und Syntaxdateien

■ Datenvergleich

- Werkzeuge zum Vergleich von Syntaxdateiversionen

■ Datenverarbeitung

- Ausführen von statistischen Berechnungen an Forschungsdaten

■ Hauptfunktionen in einer virtuellen Forschungsumgebung (2)

■ Datenanbieterzugriff

- Zugang zu Ausgangsdaten von FDZ (SUF)
- Fernrechnen und Ablage von Daten zur langfristigen Archivierung

■ Konfiguration

- System-, Installations- und Nutzer/innen Einstellungen

■ Verwaltung

- Verwaltung von Benutzergruppen, Rollen und globalen Rechten

■ Kollaboration

- Werkzeuge für die Zusammenarbeit im Verbund

■ Publikation

- Veröffentlichung von Forschungsergebnissen im Forschungsverbund sowie auf Webseiten

■ Wie die virtuelle Forschungsumgebung Arbeitsweisen von Forscher/inne/n verändert

■ Arbeit im Verbund ...

- ... bleibt dezentral und individuell
- ... wird durch technische Vernetzung transparenter
- ... setzt künftig Bereitschaft zu »syntax sharing« voraus
 - Syntax an (teil-)öffentlichem Ort

■ Verhaltensanforderungen an Verbundpartner/innen ...

- ... werden technisch »hinterlegt« und formalisiert, damit für FDZ wirksamer
- ... Nutzer/inne/n werden bei Einhaltung von Standards wirksam unterstützt

■ Verbundkoordination ...

- ... wird anspruchsvoller, um technischen Support erweitert

■ Softwareentwicklung in engem Kontakt mit Forschungspraxis

■ Kollaborative Datennutzung aus Sicht des Datenschutzes – Probleme und Lösungen (1)

■ Unproblematisch:

- Ablage von Metadaten, gemeinsamer Zugriff auf Syntaxdateien in der Forschungsumgebung

■ Probleme mit Forschungsdaten in virtuellen Organisationen:

- Personelle Nutzungsbeschränkung, Genehmigungsvorbehalt?
- Inhaltliche Nutzungsbeschränkung (Zweckbindung)?
- Zeitliche Nutzungsbeschränkung (Befristung, Löschrfristen)?

■ Mögliche technische Lösungen in virtueller Arbeitsumgebung

- Zugriffskontrolle durch Rechteverwaltung für virtuelle Organisationen: zertifikatsbasierte Authentifizierung (Grid Security Infrastructure: PKI)
- Transparenz der Datennutzung durch gemeinsame Dokumentation des »Lifecycles« (Protokollierung, Historie)
- Rechteverwaltung gewährleistet rollenbasierte Rechte und Löschrfristen

■ Kollaborative Datennutzung aus Sicht des Datenschutzes – Probleme und Lösungen (2)

■ Derzeitige Nutzungsverträge unterstellen ...

- verantwortliche (aufsichtspflichtige, haftende) Vertragspartner
- Einhaltung des Datensicherheitskonzepts in hierarchischer Organisation
- Datenablage auf sicheren Datenträgern beim jeweiligen Vertragspartner – keine Datenübermittlung

■ Mögliche Lösungen für virtuelle Organisation

- Ausschließlich berechtigte Nutzer/innen mit Einzelnutzungsverträgen
- Übernahme geprüfter Datensicherheitskonzepte durch alle Partner/innen
- Keine Ablage, Verarbeitung von Forschungsdaten in virtueller Umgebung
- Physikalische Datenablage in Repository bei einer beteiligten Forschungseinrichtung (Service Level Agreement)
- ISO-Normen für Datensicherheit, sichere Leitungen
- FDZ in Forschungsverbund und virtuelle Organisation einbeziehen

■ Kollaborative Datennutzung aus Sicht des Datenschutzes – Probleme und Lösungen (3)

■ **Nachteil: Datenaustauschplattform wird nur teilweise genutzt**

- Beschränkung der Kollaboration auf Syntax und Outputs
- Keine gemeinsame Langzeitarchivierung von Forschungsdaten

■ **Über den SUF hinaus – Remote Access in virtueller Arbeitsumgebung**

- Daten werden in Forschungsdatenzentren verarbeitet
- Datenfernverarbeitung wird durch Plattform technisch unterstützt
- Kollaboration auch zu schwach anonymisierten Daten möglich
- Langzeitarchivierung von Forschungsdaten bei FDZ

■ Gemeinsame Interessen von FDZ und Virtuellen Forschungsumgebungen?

■ Rechtliche Klärung

- Datensicherheitskonzepte für Datenaustauschplattformen

■ Technische und rechtliche Entwicklungen bei Remote Access

- Lösungen zur sicheren Datenfernverarbeitung
- Vernetzung beim Datenzugang (»FDZ in FDZ«)

■ Datennutzung und Datenqualität

- Retrieval-fähige, reanalysierbare Langzeitarchivierung von Auswertungen
- Verknüpfung von Datenbereitstellung und Forschung – auch in FDZ
- Syntax-Sharing, nachnutzbare generierte Variablen
- Gemeinsame Metadatenstandards von Datenhaltern und Nutzer/innen
- Sicherung guter wissenschaftlicher Praxis durch Dokumentation
- Mehr Rückmeldungen aktiver Nutzer/innen zu Daten

■ Dokumentation

■ **Berichterstattung zur sozioökonomischen Entwicklung Deutschlands**

- <http://www.soeb.de/virtaug.php>

■ **WissGrid – Grid für die Wissenschaft**

- http://www.wissgrid.de/publikationen_de.html