

Ergebnisse des 4. Workshops zum Modellprojekt VirtAUG

Auswirkungen einer Virtuellen Arbeitsumgebung auf Arbeitsabläufe in der sozioökonomischen Forschung und Berichterstattung am 15.12.2010 im SOFI, Göttingen

Teilnehmer/innen

Mara Boehle, GESIS, Mannheim

Dr. Thomas Drosdowski, GWS, Osnabrück

Dr. Harry Enke, WissGRID, AIP, Potsdam

Christian Gerhards, Universität Bielefeld

Patrick Harms, WissGRID, SUB Göttingen

Dr. Florian Köhler, FDZ LSKN Niedersachsen, Hannover

Jens Ludwig, WissGRID, SUB Göttingen

Hans Nerlich, PT DLR, Bonn

Ewa Sojka, INIFES, Stadtbergen

Prof. Dr. Christoph Weischer, Universität Münster

PD Dr. Jürgen Kädtler, SOFI, Göttingen

Dr. Peter Bartelheimer, Tanja Schmidt, René Büttner, Sarah Cronjäger (SOFI)

Tagesordnung

- Begrüßung, Einführung: Peter Bartelheimer, SOFI
- Präsentation „Arbeiten im Verbund – nichttechnische und technische Anforderungen“:
Peter Bartelheimer, SOFI
- Präsentation „Ein Tag in der virtuellen Arbeitsumgebung – Architekturskizze und Anwendungsfall“: Tanja Schmidt, SOFI
- Themenanmeldungen für die Diskussion
- Diskussionsrunden zu den Funktionalitäten:
 - Konfigurations- und Verwaltungswerkzeuge (u.a. Rechteverwaltung)
 - Datenbezogene Arbeitswerkzeuge
 - Werkzeuge für die Ausbaustufe (Kollaborations-, Datenanbieter- und Publikationswerkzeuge)
- Ergebniszusammenführung und -sicherung, Abschlussdiskussion

Diskussionsergebnisse

Auf Basis der Präsentationen ergaben sich durch Kartenabfrage verschiedene Diskussions-Cluster.

Cluster 1: Virtuelle Arbeitsumgebung (VAU)

Es handelt sich dabei um keine fertige Anwendung, sondern um ein Komponentensystem mit Modulen, der Zugang geschieht via Browser über ein Portal, existierende Komponenten liegen gebündelt beim Provider und es werden möglichst kleine Pakete gepackt. Die Module können über eine Benutzeroberfläche an- und abgewählt werden.

Das System ist absturzsicher: Es wird „backend“ durch professionellen Provider und überinstitutionelle Ressourcen sichergestellt. Dementsprechend ist zwar möglich, dass Einzelkomponenten abstürzen, aber die Daten wären nicht weg. Die Implementation der VAU hat noch nicht begonnen.

Welche Instrumente sind sinnvoll für die Zusammenarbeit? Das Instrument der Versionsverwaltung, kann als kollaborative Komponente gesehen werden, denn jede/r Projektpartner kennt und kann auf den letzten Stand der anderen Projektpartner/innen zugreifen. Nutzer/innen loggen sich auf ihrem Desktop in das Eingangsportal ein um die Komponenten nutzen zu können. Dabei können parallel auch lokale Programmkomponenten genutzt werden (z.B. auch in TextGrid).

Der zentrale Server steht bei einem überregionalen Provider wie der GDWG; bei diesen sind dann professionelle Leistungen einzukaufen.

Cluster 2: Datenschnittstelle

Es gibt nicht nur Probleme, sondern auch Lösungen – interessant für die gesamte Datenlandschaft.

Die FDZ haben Interesse an der Zertifikats-Infrastruktur für remote access und auch an aktiver Forschungsbeteiligung.

Cluster 3, andere, aggregierte Daten

Wo enden Beschränkungen aufgrund des Datenschutzes? Das heißt, wann findet der Übergang zu öffentlichen Daten in der VAU statt? Diese Frage stellt sich besonders bei Betriebsdaten – wer hat die Rechte an diesen Daten? Insbesondere bei Organisationsdaten von Betriebsfallstudien wurde für den Betriebszugang Anonymität vorausgesetzt. Daraus entstehen Identifikationsprobleme. Die VAU setzt damit ein Umfeld voraus, das durch die Datenhalter definiert wird. Neue „Umgebungsprobleme“ entstehen dann durch andere, weitere Datenhalter, z.B. Betriebsdaten.

Urheber- und Nutzungsrechte: öffentlich finanzierte Daten sind für die Wissenschaft frei.

Aggregatdaten bedürfen einer detaillierten Beschreibung, was einen großen Aufwand bedeutet. Und insbesondere Organisationsdaten haben datenschutztechnisch spezielle Anforderungen und andere „Umgebungsprobleme“.

Metadaten sollen nicht nur technische Entstehungsinformationen enthalten, sondern müssen die Daten auch fachwissenschaftlich charakterisieren.

VAU muss einerseits den Datenzugang für geschlossene Arbeitsgruppen unterstützen, andererseits aber auch die Arbeit in gemeinsam benutzten Bereichen. Es muss ein Wissen bestehen, welche Daten es im Projekt sonst noch gibt. Die Möglichkeit, sich zu informieren, kann nur technisch unterstützt werden (Datei- und Literaturverwaltung). Eventuell können die Nutzer/innen auch durch Meldungen aus dem System hingewiesen oder erinnert werden. Jedoch bleibt die Frage offen, wer das moderiert. Die VAU kann auf jeden Fall die Transparenz in arbeitsteiligen Zusammenhängen erhöhen, jedoch ist dies nicht rein technisch lösbar.

Denn die Technik führt auch zu neuen Problemen. Dabei ist der Koordinationsaufwand mit IT-Problemen zusammen zu sehen. Jedoch werden dadurch, dass technische Umsetzungen notwendig sind, Organisationsanforderungen bewusster, und der Koordinationsaufwand wird sichtbarer.

Insgesamt betrachtet befördern die technischen Abläufe kulturellen Wandel im Arbeiten und Publizieren, es gibt kaum noch Einzelautoren.

Welche bisher bereits formalisierte Teilarbeitsprozesse können durch die neue technische Basis gewinnen? VAU ist besonders da effektiv, wo vorher schon viel Standardisierung vorlag.

Cluster 4, Metadaten

Der DDI-Standard sollte bei Erstellung von Metadaten nicht in ganzem Umfang angewendet werden. DDI-Datensätze von externen Daten können auch importiert und nachgenutzt werden. Andererseits können selbst erstellte Metadaten im DDI-Format für andere Anwendungen exportiert werden. Problematisch ist nur, festzulegen, welche Informationen in die Metadaten eingehen sollen. Wenn man zu viel von den Nutzenden verlangt, bekommt man nichts. Evtl. Schlagworte vorgeben? Man wird sich dauerhaft damit beschäftigen müssen. Möglich wäre, auf die bereits aufbereiteten Metadaten von definierten Datensätzen der FDZ, z.B. Mikrozensus (auch über Microdata Lab) via links zuzugreifen und diese gezielt durchzuarbeiten. Außerdem lässt sich Prozesswissen nicht über Daten formalisieren.

Metadatenysteme sind nie fertig; Felder werden „missbraucht“, mehrere Informationen werden in ein Feld eingetragen.

Welche Metadaten werden standardmäßig abgefragt? Welche technischen und fachlichen Daten werden extrahiert? „Kommentarfelder auszählen“.

Wer bekommt die Lizenz zum Löschen?

Cluster 5 Syntax Sharing

Es ist zu unterscheiden, ob iterativ oder parallel an Syntax gearbeitet wird. Bei beiden Anwendungsfällen treten Koordinationsprobleme auf, die nicht rein technisch zu lösen sind.

In der VAU sollten so viele generierte Variablen wie möglich verfügbar sein – eventuell nicht nur als Ergebnis aus dem Arbeitsprozess, sondern auch als Datenservice-Funktion. Hier ist die Servicequalität der VAU gefragt, z. B. wenn sich Variablennamen im SOEP ändern?

[Schlussdiskussion]

Virtuelle Forschungsumgebungen stoßen auf Akzeptanzprobleme in Sozial- und Geisteswissenschaften. Es sollte aufgezeigt werden, dass sie ein nützliches Instrument sind. Gegenstand einer neuen Bekanntmachung des BMBF (eHumanities, bis 30.09.2011) sind sowohl Infrastrukturprojekte als auch Forschungsprojekte.

Ideal wären Anwendungsfälle, wie in den Textwissenschaften, in denen durch VAU Fragen beantwortet werden können, die vorher nicht zu beantworten waren, und noch besser, die vorher gar nicht gestellt wurden.

Eine VAU ist sinnvoll als eine Möglichkeit für bestimmte komplexe Projekte, ihre Nutzung sollte jedoch keine Verpflichtung, sondern freiwillig sein.

Der verbesserte Datenzugang wäre ein wesentlicher Vorteil, den eine VAU allen sozialwissenschaftlichen Projekten bietet. Sonst ist eine VAU nur sinnvoll für Verbundprojekte. Darüber hinaus könnte sie Serviceleistungen für einzelne Datensätze (Modell: MISSY) oder Datenmanagement im Zeitverlauf unterstützen.

Durch die VAU wird die Schnittstelle zu den Daten von der „Verbraucherseite“ her gestärkt. Durch paralleles überörtliches Arbeiten lokal verteilter Wissenschaftler/innen nähern sich zwei unterschiedliche Akteursgruppen an, nämlich die Datenbesitzer/innen und die Datennutzer/innen.

Zusammenfassend ist immer zu bedenken, dass es nicht allein um technische Möglichkeiten geht, sondern um eine gemeinsame sozialwissenschaftliche „Story“.