

Ergebnisse des 3. Workshops zum Modellprojekt VirtAug

Virtuelle Arbeitsumgebung für sozioökonomische Forschung und Berichterstattung –
Rechtliche Aspekte der Nutzung von Forschungsdaten. Göttingen, 05. November 2010

Teilnehmer/innen

Stefan Bender, FDZ BA, IAB, Nürnberg

Tatjana Mika, FDZ RV, Berlin

Michael Schmaus, FDZ der Statistischen Landesämter, Düsseldorf

Dr. Jürgen Schupp, DIW, Berlin

Dr. Michael Stegmann, FDZ RV, Würzburg

Dr. Heike Wirth, GESIS, Mannheim

Hans Nerlich, PT DLR, Bonn

Dr. Harry Enke, WissGRID, AIP, Potsdam

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

Zur Expertise und zu den Präsentationen

Der Begriff „Originaldaten“ sollte nicht für SUF verwendet werden, da es sich um anonymisierte Daten handelt. Einheitlich sollte in den Darstellungen von Ausgangsdaten gesprochen werden. Zudem sollte unterschieden werden zwischen „personenbezogen“ versus „Personen beziehend“.

Welchen Sicherheitsstandards Provider von Speicherplatz bieten, ist nicht generell durch das GRID festgelegt. Provider einer virtuellen Arbeitsumgebung sollten ein account management für virtuelle Organisationen bieten.

Erster Diskussionsblock: Vertragsgestaltung zur Datenbereitstellung, Datenschutzkonzepte

Die Vertreter/innen der an der Diskussion beteiligten FDZ zeigten generelle Bereitschaft, an der Entwicklung mitzuwirken, verwiesen aber auf aus ihrer Sicht zentrale Datenschutzprobleme. Sie stellten übereinstimmend fest, dass für alle Forschungseinrichtungen, die in einer virtuellen Arbeitsumgebung (VAU) kooperieren, Einzelnutzungsverträge mit den (berechtigten) Dateneinrichtungen bestehen müssen. Die Vertragsgestaltung setzt verantwortliche (aufsichtspflichtige) Antragsteller voraus, die in einer hierarchischen Arbeitsorganisation unterbinden können, dass Mitarbeiter/innen Datenschutzanforderungen verletzen, und die eine Nutzung ausschließlich für den vertraglich bestimmten Forschungszweck an lokalen Standorten sicherstellen. Die Verträge der amtlichen FDZ die Zahl und die Identität der Nutzer/innen fest; auch andere FDZ verlangen die Benennung der Stellen und Personen, die

personenbezogene Daten bearbeiten. Die virtuelle Organisation kommt als Vertragspartner/in nicht in Betracht. Gruppen- oder Verbundverträge können auch für die beteiligten Nutzer/innen ein erhöhtes Haftungsrisiko bedeuten: Bei einem individuellen Verstoß gegen den Datenschutz (z.B. einer unzulässigen Datenverknüpfung) würden alle Verbundeinrichtungen von der Datennutzung ausgeschlossen.

Ein FDZ wägt ab, dass eine VAU einerseits die vertraglich unterstellte Hierarchie schwächt, andererseits aber einen weniger privaten Raum für die Datennutzung schafft (Syntax liegt an einem öffentlichen Ort). Ein anderes FDZ kann sich vorstellen, dass SUF in einer VAU gehalten werden und laufen, wenn diese gegen nicht nutzungsberechtigte Dritte abgegrenzt ist (nicht jedoch z.B. regionale Kontextdaten, die das Risiko der De-Anonymisierung erhöhen würden. Oder der Provider wäre selbst eine Forschungseinrichtung – in der Diskussion blieb offen, ob das z.B. auf die GWVG zuträfe – und als solche Verbundpartner.

Konsens besteht darüber, dass der „Idealzustand“ für eine VAU Datenfernverarbeitung (remote access processing) über eine remote-access-Plattform wäre. Die Daten würden das FDZ gar nicht verlassen, damit ist auch die Gefahr durch unsichere Leitungen gebannt. Zukunftsträchtig wäre eine gemeinsame Remote-Access-Architektur (nächster Schritt „FDZ in FDZ, Zugangskontrolle ohne Verletzung von Persönlichkeitsrechten). In dieser Richtung zeichnen sich Parallelentwicklungen ab (IAB-Ausschreibung Metadatenbanksystem, EU-Projekt Data Without Boundaries, DDI-Datendokumentation).

Ein weiteres rechtliches Problem liegt darin, ob bestehende Datenschutzkonzepte beteiligter Forschungseinrichtungen auf eine VAU übertragbar sind. Ist dies nicht der Fall, müsste das Sicherheitskonzept für die Datenaustauschplattform von den FDZ, die Daten bereitstellen sollen, zeitaufwändig und mit unsicherem Ergebnis geprüft werden.

Zweiter Diskussionsblock: Nutzungsbeschränkungen, Auswertungsprogramme, Metadaten

Aus Sicht der FDZ sind für die Lösung von Fragen der personellen Nutzungsbeschränkung, Befristung und Zweckbindung zwei Szenarien zu unterscheiden: Liegen die Forschungsdaten nur bei einem zentralen Provider oder auch auf individuellen Workstations. Was ist mit Arbeitsdateien? Wann dürfen Mikrodaten auf die Workstation gespeichert werden? Lässt sich abgleichen, was auf dem Repository, was auf Workstations vorhanden ist? Kann man einen gemeinsamen Ort des Datenaustauschs in individuelle Datennutzungsverträge mit Partnereinrichtungen einer VAU aufnehmen? Was für die Nutzer/innen am einfachsten wäre, ist datenschutzrechtlich am kompliziertesten.

Partnereinrichtungen einer VAU müssten als berechtigte Nutzer/in anerkannt sein. Mit dem Provider einer VAU, der nicht zur scientific community gehört, hätten die FDZ kein Vertragsverhältnis. Forscht der Provider nicht selbst und lägen die Daten bei ihm, betriebe er eine rechtlich nicht zulässige Datenvorrathaltung. Eine Lösung könnte darin bestehen, dass Provider Vertragspartner werden – etwa ein Rechenzentrum als An-Institut, so dass ggf. Vertrag mit der Universität geschlossen werden könnte („vertragstechnisch am schön-

ten“). Alternativ wären die Provider (z.B. ein Rechenzentrum) zu zertifizieren. Die Zertifizierung müsste sicherstellen, dass Mitarbeiter/innen des Rechenzentrums die Daten nicht einsehen können und ISO-Normen für Datensicherheit eingehalten werden; zu den Anforderungen könnten auch sichere Leitungen gehören.

Lösungen für schwach anonymisierte Daten wären noch komplizierter.

Für das Verbot der Datenzusammenführung würde die Datenhaltung auf einem Rechner kein neues Problem darstellen.

Personelle Nutzungsbeschränkung und Genehmigungsvorbehalt wären in der skizzierten Architektur gewährleistet: Man kann gruppenbasierte Personenrechte nachvollziehen. Die Protokollierung durch eine dritte Instanz erlaubt sogar eine bessere Nutzer/innenkontrolle – eine Qualität, die man gegenüber den FDZ herausstellen sollte.

Für die FDZ ist die Befristung der Nutzung entscheidend. In einem größeren Projekt sollte die Nutzung für vier oder fünf Jahre beantragt werden. Überlässt die bisherige Vertragsgestaltung die tatsächliche Entscheidung, wann welche Daten gelöscht werden, den Nutzer/inn/en überlässt, wäre dies in einer VA dies technisch gesichert und nachvollziehbar.

Einige FDZ bieten den Nutzer/inn/en nach Ablauf der Nutzungsfrist für Arbeitsdateien Langzeitarchivierung (zehn bis 20 Jahre) und Wiederaufrufbarkeit. Dabei sichert die von den Nutzer/innen entwickelte Syntax die Nachnutzbarkeit; Programme müssen auf den Ursprungsdatensätzen lauffähig sein. Die FDZ verweisen auf ihre Kompetenz bei Datenmanagement, Versionierung und Versionskontrolle (einzigartig zitierbar über DOI, Persistent Identifier).

Ein offener Diskussionspunkt blieb, ob eine Archivierung beim Bundesdatenarchiv eine weitere Alternative zur individuellen Archivierung darstellt.

Verwiesen wurde darauf, dass in der amtlichen Statistik die Nutzungsgebühren erhöht werden, da die FDZ sich teilweise über Gebühren finanzieren müssen, was auch Veränderungen für die möglichen Nutzungsfristen nach sich zieht.

Metadaten und Auswertungssyntax unterliegen keinen Vertragsbeschränkungen. Bei SUF haben Nutzer/innen völlige Freiheit bei Programmierung und Dokumentation; für Fernrechnen und Onsite-Nutzung bestehen Anforderungen an die Dokumentation. Die FDZ beklagen, dass an Informationen über generierte Variablen, Wissen über Datenmängel und Datenqualität bislang kaum etwas an die Daten haltenden Einrichtungen zurückkommt. Die FDZ bekommen auch die vertraglich zustehenden Publikationen nicht. Metadaten-systeme mit Wiki-Elementen setzen generell eine andere Bereitschaft der Nutzer voraus, sich an „kollektiven Gütern“ zu beteiligen.