

## **Gemeinsamer Sach- und Schlussbericht**

für das Verbundprojekt

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

(Stand: 13.10.2014)

**Zuwendungsempfänger:** Soziologisches Forschungsinstitut Göttingen (SOFI) e.V.

Förderkennzeichen: 01UW1203A

Laufzeit des Vorhabens: 01.06.2012 – 30.06.2014

Berichtszeitraum: 01.06.2012 – 30.06.2014

Ansprechperson:

Dr. Peter Bartelheimer

Soziologisches Forschungsinstitut Göttingen (SOFI)

Friedländer Weg 31, D-37085 Göttingen

Tel. +49 551 52205-51; Fax +49 551 5220588

email: peter.bartelheimer@sofi.uni-goettingen.de

Der nachstehende Sach- und Schlussbericht wird von den Partnereinrichtungen des Verbundvorhabens „Virtuelle Forschungsumgebung für die Sozioökonomische Berichterstattung (VFU soeb 3)“ gemeinsam vorgelegt.

**Zuwendungsempfänger:** **Soziologisches Forschungsinstitut Göttingen (SOFI) e.V.**

Förderkennzeichen: 01UW1203A

Unterauftragnehmer:

Dr. Tanja Schmidt – Sozialforschung

Janine Lange/ Christoph Rechenberg (Forschungsdatenzentrum der Rentenversicherung)

Dr. Jan Goebel (Forschungsdatenzentrum SOEP)

**Zuwendungsempfänger:** **Gesellschaft für Wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)**

Förderkennzeichen: 01UW1203B

**Zuwendungsempfänger:** **GESIS - Leibniz-Institut für Sozialwissenschaften e.V.**

Förderkennzeichen: 01UW1203C

**Zuwendungsempfänger:** **Georg-August-Universität Göttingen/ Durchführende Einrichtung: Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB)**

Förderkennzeichen: 01UW1203D

Unterauftragnehmer: Pro. Dr. Uwe Schwiegelshohn (D-Grid gGmbH)

**Zuwendungsempfänger:** **Bundesagentur für Arbeit (BA) - Institut für Arbeitsmarkt- und Berufsforschung**

Förderkennzeichen: 01UW1203E

## Inhaltsverzeichnis

<b>I. Kurzdarstellung .....</b>	<b>5</b>
I.1 Aufgabenstellung .....	5
I.2 Voraussetzungen .....	6
I.3 Planung und Ablauf.....	9
I.4 Wissenschaftlicher und technischer Stand .....	11
I.5 Zusammenarbeit .....	13
<b>II. Eingehende Darstellung.....</b>	<b>14</b>
II.1 Erzielte Ergebnisse .....	14
II.1.1 Arbeitspaket 1: „Projektkoordination, Projektmanagement“ (SOFI).....	14
II.1.2 Arbeitspaket 2: Fachwissenschaftliche Anforderungen, Anwendungsfälle (SOFI).15	15
II.1.3 Arbeitspaket 3: „Gesamtarchitektur“ (SUB).....	19
II.1.4 Arbeitspaket 4: „Forschungsdaten“ (FDZ IAB, FDZ SOEP, FDZ RV).....	21
II.1.5 Arbeitspaket 5: „Portal und Hosting“ (GWDG) .....	25
II.1.6 Arbeitspaket 6: „Middleware“ (GWDG) .....	27
II.1.7 Arbeitspaket 7: „Metadaten“ (GESIS) .....	29
II.1.8 Arbeitspaket 8: „Kollaborative Werkzeuge“ (GESIS).....	30
II.2 Zahlenmäßiger Nachweis .....	31
II.3 Notwendigkeit und Angemessenheit der geleisteten Arbeit.....	32
II.4 Voraussichtlicher Nutzen .....	32
II.5 Fortschritt bei anderen Stellen .....	33
II.6 Veröffentlichungen und Transfer .....	33
II.6.1 Veröffentlichungen.....	33
II.6.2 Vorträge .....	34
II.7 Literatur .....	36

## Verzeichnis der Tabellen und Abbildungen

Abbildung 1: Mockup Ergebnisliste aus facettierter Suche.....	15
Abbildung 2: Registrierung, Mitglieder, Absprechen und Kalender der VFU.....	16
Abbildung 3: Kollaboration im Wiki und Forum.....	17
Abbildung 4: Dateisystem .....	17
Abbildung 5: Daten, Analysen und Archivsuche.....	18
Abbildung 6: Job Submission .....	24
Abbildung 7: Remote Desktop.....	24
Abbildung 8: Komponenten-Architektur der <i>VFU soeb 3</i> .....	27
Tabelle 1: Ausgaben nach Positionen im Gesamtfinanzierungsplan .....	32

## I. Kurzdarstellung

### I.1 Aufgabenstellung

Ziel des Projekts war die Entwicklung des Prototyps einer webbasierten Virtuellen Forschungsumgebung (*VFU soeb 3*), die die Zusammenarbeit räumlich verteilter wissenschaftlicher Einrichtungen in Verbundprojekten unterstützt. Zu verbessern waren insbesondere die Möglichkeiten kollaborativer Auswertung sozialwissenschaftlicher Mikrodaten über die gesamte wissenschaftliche Leistungskette, vom Datenzugang bis hin zur Archivierung von Forschungsergebnissen.

In einer Entwicklungs- und Einführungsphase von 24 Monaten entstand der Prototyp eines VFU-Portals, das mit ersten kollaborativen Werkzeugen in der Forschungspraxis erprobt wurde und mit Projektabschluss einsetzbar ist. Fachwissenschaftliche Anforderungen an eine VFU wurden spezifiziert, und mit dem entwickelten Prototyp wurden Nutzungsstudien durchgeführt. Vorrangige Entwicklungsziele waren Werkzeuge zur Entwicklung, Dokumentation, Archivierung und Nachnutzung von Syntax für Statistikprogramme (Syntaxeditor) und die Entwicklung eines gemeinsamen Metadatenstandards sowie dessen technische Unterstützung (Metadaten- und Studieneditor). Weitere Entwicklungsschwerpunkte für das VFU-Portal und für die Middleware waren u.a. eine Sicherheitsarchitektur mit personalisierten Zugangsrechten, notwendige Funktionen der Informations- und Datenverwaltung und eine grafische Benutzeroberfläche für die erforderlichen Konfigurationswerkzeuge. Außerdem sollte die Architektur für das Portal der VFU und die integrierte Middleware bereits künftige Entwicklungsmöglichkeiten beim Datenfernzugriff berücksichtigen. Durch die beteiligten FDZ wurden datenschutzrechtliche Anforderungen und technische Voraussetzungen für Microdata Access spezifiziert. Das Projekt sollte die Erfolgsaussichten für neue Zugangswege zu Mikrodaten in Dateneinrichtungen verbessern und Lösungen für deren Unterstützung in einer VFU skizzieren. Entwicklungsaufgaben für eine insbesondere um Werkzeuge für Remote Access erweiterte Version der VFU wurden spezifiziert und an einem Proof of Concept demonstriert.

Das Projekt wurde in Kooperation folgender Partnereinrichtungen durchgeführt: Soziologische Forschungsinstitut Göttingen (SOFI) e.V., GESIS – Leibniz-Institut für Sozialwissenschaften e.V. (Mannheim), Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG), Forschungsdatenzentrum (FDZ-IAB) der Bundesagentur für Arbeit (Nürnberg) im Institut für Arbeitsmarkt- und Berufsforschung (IAB), Georg-August-Universität Göttingen Stiftung Öffentlichen Rechts (ausführende Stelle: Niedersächsische Staats- und Universitätsbibliothek (SUB)), Forschungsdatenzentren der Rentenversicherung (FDZ-RV) und FDZ sozio-oekonomisches Panel (FDZ-SOEP), D-Grid gGmbH.

## I.2 Voraussetzungen

Zwei Entwicklungen haben die Forschungsinfrastruktur für die Sozialwissenschaften in den letzten zehn Jahren rasch und grundlegend verändert. Zum einen entstanden in Umsetzung der 2001 vorgelegten Empfehlungen der Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI) Forschungsdatenzentren und Datenserviceeinrichtungen, die immer mehr amtliche Daten und öffentlich finanzierte Umfragedaten als Mikrodaten für Forschungszwecke zugänglich machen. Zum anderen verändern die technische Entwicklung und zunehmende Verfügbarkeit netzbasierter IT-Ressourcen und IT-Dienste den wissenschaftlichen Zugang zu Daten, die Datenbereitstellung und die datenbasierte Forschungsarbeit.

„E-Science“ oder „e-Infrastruktur“ sind gängige Bezeichnungen für diese sowohl daten- als auch technikgetriebenen Veränderungen. Grundsätzlich haben die damit etikettierten technischen Innovationen das Potenzial, die gesamte „Leistungskette“ der wissenschaftlichen Arbeit zu verändern: Die digitale Vernetzung räumlich verteilter Arbeitsplätze, von denen aus über „Middleware“ (d.h. über Programme, die zwischen Anwendungen vermitteln) und Dienstprogramme auf Rechnerkapazitäten und Daten zugegriffen werden kann, soll neue standortunabhängige Formen der Zusammenarbeit in der Forschung ermöglichen und unterstützen.

Häufig wird der Aufbau einer eigenen Netz-Infrastruktur als eine Voraussetzung für „e-Science“ angesehen. Seit 2004 förderte das BMBF mit der deutschen „D-Grid-Initiative“ eine Erweiterung des Internets durch Hochleistungsnetzwerke („Grids“), die der Wissenschaft „nahezu unendlich große Rechen- und Speicherkapazität, Flexibilität und automatische Anpassung von komplexen Rechenprozessen durch dynamischen und konzertierten Betrieb der vernetzten Ressourcen, höhere Qualität der Ergebnisse durch Grid-unterstützte Entwicklung und schließlich Einsparungen durch eine verbrauchsorientierte Abrechnung“ versprechen (Gentzsch 2007: 9). Zugleich sollen die Potentiale des „Höchstleistungsrechnens“ auch für Unternehmen erschlossen werden. Solche „On-Demand-Infrastrukturen“ werden in den letzten Jahren auch unter der Bezeichnung „Cloud Computing“ entwickelt und diskutiert.

In der D-Grid-Initiative waren überwiegend naturwissenschaftliche Fach-disziplinen vertreten, die einen hohen Bedarf an großen Rechnerleistungen haben, sowie mit TextGrid ein geisteswissenschaftlicher Forschungsverbund. Dagegen haben die sozialwissenschaftlichen Disziplinen bislang in der Regel keine Notwendigkeit gesehen, bei der Entwicklung von „E-Infrastrukturen“ auf Grid- oder Cloud-Computing zurückzugreifen, da ihre Daten- und Rechneranforderungen für virtuell vernetztes, standortunabhängiges Arbeiten auch im Rahmen der bestehenden Internet-Infrastruktur (dem sogenannten „Web 2.0“) erfüllt werden können. „Results of these developments are possibly less perfect than those designed for Grid appli-

cations, but they are facilitated by cooperative approaches within the science community and they take usually much less time to implement.“ (Mochmann 2010: 270) Daher wird die Weiterentwicklung der Informationsinfrastruktur für die Sozialwissenschaften in den letzten Jahren unabhängig von den technischen Lösungen für digitale Netze, also von der Entwicklung des Grid- oder Cloud-Computings diskutiert.

Nach der Definition der Allianz-Initiative ist eine Virtuelle Forschungsumgebung „eine Arbeitsplattform, die eine kooperative Forschungstätigkeit durch mehrere Wissenschaftler an unterschiedlichen Orten zu gleicher Zeit ohne Einschränkungen ermöglicht. Inhaltlich unterstützt sie potenziell den gesamten Forschungsprozess – von der Erhebung, der Diskussion und weiteren Bearbeitung der Daten bis zur Publikation der Ergebnisse –, während sie technologisch vor allem auf Softwarediensten und Kommunikationsnetzwerken basiert.“<sup>1</sup> Die Architektur einer VFU und ihre Funktionen lassen sich nicht standardisieren. So empfiehlt die Kommission Zukunft der Informationsinfrastruktur (KZII), dass VFU ausdrücklich die fachlich-inhaltliche Vielfalt unterstützen und „je nach fachspezifischen und individuellen Charakteristika strukturell sehr weit ausdifferenziert sein“ sollen (KZII 2011: B74). Die KZII geht von einem Bedarf von wenigstens fünf VFU für jedes der 48 Fachkollegien der DFG aus. Sicher gilt generell, dass die Arbeitsweisen der Wissenschaftler/innen Grundlage für die zu entwickelnden Informationsangebote und Dienstleistungen sein sollen (Winkler-Nees 2011). Doch VFU „nehmen stärkeren Bezug auf den tatsächlichen Gebrauch von IT-Diensten in den jeweiligen Disziplinen oder interdisziplinären Feldern, den dort eingesetzten Methoden und zu bearbeitenden Forschungsfragen“ (DFG 2010: 25). Aus beiden Gründen können VFU nur „forschungsnah“ entstehen – ihre Entwicklung bedeutet zugleich eine technische und eine soziale Innovation. Nach Auffassung der KZII werden „Virtuelle Forschungsumgebungen (...) nur dann intensiv genutzt, wenn sie die Anforderungen der jeweiligen Forschungsprozesse möglichst gut abdecken. Deshalb können die Umgebungen nicht mehr wie bisher oftmals für die Fachwissenschaftler, sondern nur mit ihnen gemeinsam entwickelt werden. Das heißt, dass Forschende den Aufbau und die Weiterentwicklung von virtuellen Forschungsumgebungen aktiv mitgestalten und mit steuern: Die kooperative Entwicklung unter Beteiligung von Forschern, Informationsspezialisten und IT-Fachleuten gehört zu den entscheidenden Erfolgsfaktoren einer virtuellen Forschungsumgebung. Die Fachwissenschaftler formulieren ihre Anforderungen und begleiten die informationelle und softwaretechnische Entwicklung der virtuellen Forschungsumgebungen in einem iterativen Prozess.“ (KZII 2011: B87f.)

Eine Bestandsaufnahme der Deutschen Forschungsgemeinschaft vom Dezember 2013 zeigt, dass im Förderbereich Informationsmanagement der DFG bislang Archivportale, Do-

---

<sup>1</sup> <http://www.allianzinitiative.de/de/handlungsfelder/virtuelle-forschungsumgebung.html>, Zugriff am 07.10.2014.

kumentationssysteme und Infrastrukturen gefördert wurden (DFG 2013). Entwicklungsarbeiten zu VFUen für die quantitativ-empirisch arbeitenden Sozial- und Wirtschaftswissenschaften sind bislang nur als INF-Projekte einzelner Sonderforschungsbereiche bekannt (vgl. z.B. Friedhoff u.a. 2013 für den SFB 882). Das DFG-Projekt Rahmenbedingungen einer disziplinübergreifenden Forschungsdateninfrastruktur zeigte den Bedarf, den Zugang zu sensitiven Daten bei datenspeichernden Einrichtung zu verbessern und in den Workflow des Forschungsdatenmanagements zu integrieren (RADIESCHEN 2013: 22 ff.).

Das Projekt „Virtuelle Forschungsumgebung für die Sozioökonomische Berichterstattung (VFU soeb 3)“ hat die Empfehlungen der KZII berücksichtigt und die VFU forschungsnah und projektspezifisch entwickelt. Da die Entwicklung einer „generischen“ Virtuellen Forschungsumgebung für die quantitativ-empirisch arbeitenden Sozialwissenschaften ohne den Anwendungsbezug konkreter Forschungsvorhaben nicht möglich gewesen wäre, wurden Infrastrukturentwicklung und fachwissenschaftliche Forschung auf neue Weise projektförmig verknüpft: Das vom BMBF geförderte dritte Verbundvorhaben „Berichterstattung zur sozioökonomischen Entwicklung Deutschlands“ (soeb 3) diente der projektbezogenen, forschungsnahen Entwicklung der VFU als Anwendungsfall. Ziel war, über den Forschungsverbund soeb 3 hinaus Lösungen von allgemeiner Bedeutung für die quantitativ-empirisch forschenden Sozialwissenschaften zu erarbeiten. In diesem Bereich gab es bislang keine Entwicklungen zu VFU, die, wie oben bereits erwähnt, vorrangig in naturwissenschaftlichen und geisteswissenschaftlichen Disziplinen stattfinden (vgl. auch Neuroth/Lohmeier/Smith 2010).

Das Projekt sollte aber entsprechend der Zielsetzung der DFG (2013: 2 f.) zugleich ein praxistaugliches Resultat erzielen, das von einschlägigen Einrichtungen nachgenutzt werden kann. Die bisher festgestellte Beschränkung auf Lösungen für individuelle Projektziele sollte überwunden werden. Ziel war daher die Entwicklung eines VFU-Prototyps, der typische Workflows quantitativ-empirischer Sozial- und Wirtschaftswissenschaften unterstützt und eine ganze „Community“ potenzieller Nutzer/innen ansprechen kann.

Es war ebenfalls ein Projektziel, ausgewählte Dateneinrichtungen von vornherein als Partner/innen in das Infrastrukturprojekt einzubeziehen. Da diese Partner/innen ebenfalls mit eigenen Forschungsbeiträgen am Verbundvorhaben soeb 3 beteiligt sind, konnten sie die Entwicklung der VFU sowohl aus der Perspektive der Datenbereitstellung als auch aus der Nutzungsperspektive bewerten und unterstützen.

Weitere Voraussetzungen ergaben sich für das Projekt aufgrund von konzeptionellen Vorarbeiten aus dem Teilprojekt „Kollaborative Datenauswertung und Virtuelle Arbeitsumgebung“ im Rahmen des vom BMBF geförderten Projekts „Dritter Bericht zur sozioökonomischen Entwicklung Deutschlands – Durchführung einer Konzeptphase“ (August 2009 bis Dezember 2010; vgl. Bartelheimer/Schmidt 2011). Hier wurden eine Reihe fachlicher Anforderungen



spezifiziert, die sich den zentralen Funktionsbereichen der VFU zuordnen lassen: Portal und Middleware, Kollaborationswerkzeuge und Datenschnittstelle.

Anforderungen an das Portal der VFU betrafen die Datensicherheit, Datenablage, -verwaltung und -bearbeitung sowie die Konfiguration der Benutzerschnittstelle. Es erforderte eine Sicherheitsarchitektur mit personalisierten Zugangsrechten. Die Organisation, Ablage und Verwaltung von Daten sollte dem Dateibrowser eines Betriebssystems mit den üblichen bekannten Funktionen (Datei laden, Datei speichern usw.) sehr ähnlich sein. Die Datenverwaltung sollte schwerpunktmäßig die Archivierung, Dokumentation und Nachnutzung bzw. kollaborative Nutzung von Auswertungssyntax unterstützen. Neben Syntaxdateien sollten über sie auch alle anderen Dateiartern, etwa Outputdateien, Tabellen- oder Textdokumente, formatunabhängig abgelegt und mittels Inhaltssuche wie über Metadaten auffindbar verwaltet werden können.

Für die gemeinsame Arbeit an Forschungsdaten bilden Syntaxeditoren und Metadatenverwaltung die wichtigsten Kollaborationswerkzeuge einer VFU. Der wichtigste potenzielle Vorteil, den Arbeit im Verbund bietet, liegt in der gemeinsamen Arbeit an Auswertungssyntax, etwa für generierte Variablen und in deren Austausch und Nachnutzung. Allerdings setzt ein solches „Syntax Sharing“ gemeinsame Dokumentationsstandards voraus. Diese sollten im Rahmen des Projekts entwickelt werden. Standards für eine detaillierte Dokumentation aller Arbeitsschritte im empirischen Forschungsprozess, von der Operationalisierung der zu analysierenden Konstrukte über Entscheidungen zum Analysedesign bis zur Umsetzung in Syntax, sollten auch der internen Qualitätssicherung dienen. Die Regeln guter wissenschaftlicher Praxis verlangen nicht nur die Sicherung und Aufbewahrung von Primärdaten, sondern auch die Dokumentation von Resultaten und die Reproduzierbarkeit des Wegs zum Ergebnis (DFG 1998: 8). Eine VFU kann die Einhaltung solcher Standards dadurch unterstützen, dass diese in einer durchsuchbaren gemeinsamen Arbeitsplattform hinterlegt sind, welche die notwendigen Angaben extrahiert oder von den Nutzer/innen abfragt.

Die Schnittstelle zu den Forschungsdaten bildete nicht nur eine notwendige Umweltbedingung für die VFU. Hier stellten sich eigene Probleme der Forschungspraxis, deren Lösung die VFU unterstützen sollte. Wie die Datenschnittstelle der VFU zu Forschungsdaten zu gestalten war, richtete sich nach den Bedingungen der Datenbereitstellung, was in der Projektarbeit geklärt werden sollte.

### **I.3 Planung und Ablauf**

Bei der Entwicklung der Virtuellen Forschungsumgebung kooperierte das Soziologische Forschungsinstitut (SOFI) in einem interdisziplinären Projektverbund mit Einrichtungen der IT-Infrastruktur und der Dateninfrastruktur. Das SOFI als Forschungseinrichtung, die IT-Dienst-

leister und die Dateneinrichtungen brachten ihre spezifischen Perspektiven in die Zusammenarbeit ein und waren dabei unterschiedlich eng an verschiedene Fachöffentlichkeiten angeschlossen.

Unter der Gesamtkoordination des SOFI (Arbeitspaket 1) wurden auf der sozialwissenschaftlichen Seite die fachwissenschaftlichen Anforderungen an eine VFU spezifiziert und Nutzungsstudien für die entwickelten Komponenten der VFU konzipiert und durchgeführt (Arbeitspaket 2). Auf der informationstechnologischen Seite moderierte die SUB unter Beteiligung der D-Grid gGmbH die technische Arbeitsgruppe, in der die Entwicklungsarbeit zwischen den beteiligten IT-Einrichtungen unter Einbeziehung der FDZ an einer gemeinsamen Architektur ausgerichtet wurde (Arbeitspaket 3). Die GWDG übernahm federführend die Entwicklung des VFU-Portals und des Hostings sowie die Integration der Middleware (Arbeitspakete 5 und 6), GESIS die Entwicklung des Metadatenstandards und der kollaborativen IT-Werkzeuge, insbesondere der Syntax- und Metadateneditoren (Arbeitspakete 7 und 8).

Die FDZ spezifizierten die technischen Schnittstellen des VFU-Portals zu ihren Datenbeständen und IT-Verfahren und brachten Daten- und Metadatenkompetenzen in die gemeinsame Entwicklungsarbeit ein (Arbeitspaket 4); sie standen in der Mitte zwischen fachwissenschaftlichen Anforderungen und Wissenstechnologien.

Die Projektarbeit begann im Juni 2012 mit einem dreitägigen Auftaktworkshop des gesamten Verbundes *VFU soeb 3*. Im Oktober 2012 wurden auf einem weiteren gemeinsamen Verbundtreffen erste Ergebnisse diskutiert. Ein drittes gemeinsames Treffen fand am 11. Juli 2013 in Göttingen statt. Der Entwicklungsstand des Portals und der Editoren wurde vorgeführt, Ergebnisse der ersten Nutzungsstudie präsentiert, ein Ausblick zu Remote Data Access gegeben und diskutiert und die Abschlusstagung des Projekts wurde vorbereitet.

Neben den drei genannten Treffen des gesamten Verbundes fanden in der Projektlaufzeit ca. 15 interne themenspezifische Treffen in kleineren Arbeitsgruppen statt.

Eine Lenkungsgruppe aus Vertreter/innen der Arbeitspakete (AP) hat sich ca. sechswöchentlich in Telefonkonferenzen, die von der Verbundkoordination (SOFI, AP 1) ausgerichtet wurden, besprochen. Zur Koordination der technischen Arbeitspakete organisierte die SUB in Zusammenarbeit mit D-Grid GmbH (AP 3) regelmäßig im Abstand von 2 Wochen eine Telefonkonferenz der technischen APs. Hier wurde u.a. auch die Gesamtarchitektur der VFU abgestimmt und laufend entsprechend dem Projektfortschritt überprüft und weiterentwickelt.

Im Oktober 2012 wurde die projektinterne Kommunikation auf eine erste Prototyp-Version des VFU-Portals umgestellt, deren Funktionalitäten stets verbessert wurden. Im November 2013 wurde die VFU auf dem Auftaktworkshop des Verbundprojekts *soeb 3* präsentiert. Seit

Dezember 2013 arbeitet der Forschungsverbund *soeb 3* operativ mit der VFU. Die Funktionalitäten wurden fortlaufend evaluiert und an die Bedarfe des Verbunds angepasst.

Da die Entwicklung der VFU als infrastrukturelles Begleitprojekt zum sozialwissenschaftlichen Verbundvorhaben *soeb 3* angelegt war, galt für die Beziehung zwischen beiden Projekten somit, dass die Nutzer/innen aus *soeb 3* in den Entwicklungsprozess einbezogen wurden, sobald funktionsfähige Prototypen für Komponenten der VFU verfügbar waren. Die Vorhabensbeschreibung unterstellte bei der Spezifizierung der Teilaufgaben und Meilensteine für die Entwicklungs- und Einführungsphase der VFU, dass der Forschungsverbund *soeb 3* etwa im fünften VFU-Projektmonat neu zusammengesetzt sein werde und die Arbeit aufnehmen. Aufgrund von Verzögerungen im Antragsverfahren war der Start des Forschungsverbunds *soeb 3* erst im September 2013 möglich, d.h. zehn Monate später als geplant. Die ursprüngliche Meilensteinplanung musste folglich angepasst werden.

Da sich die zeitliche Überschneidung von Infrastrukturprojekt und Forschungsverbund verzögerte, wurden die beiden geplanten Nutzungsstudien für die VFU behelfsweise mit künftigen Beteiligten des Forschungsverbunds im Juni und November 2013 durchgeführt. Diese Lösung konnte aber einen längeren operativen Probetrieb des VFU-Prototyps unter realen Arbeitsbedingungen nicht ersetzen. Da sich viele Gestaltungsanforderungen und technische Probleme erst bei intensiver operativer Nutzung und im ständigen Feedback zwischen Nutzer/innen und Entwickler/innen zeigen, wurde eine Verlängerung der Projektlaufzeit um vier Monate mit dem BMBF vereinbart, um die Stabilität und Qualität des entwickelten Prototyps zu sichern und das Vorhaben erfolgreich abschließen zu können.

Die Abschlussveranstaltung des Projekts *VFU soeb 3* fand am 24. Januar 2014 in Berlin statt. Der Projektverbund präsentierte die erste operativ genutzte Version der VFU einer breiteren Fachöffentlichkeit und stellte sie zur Diskussion. Die Tagung wurde auf der Projektwebseite dokumentiert.<sup>2</sup>

#### **I.4 Wissenschaftlicher und technischer Stand**

Der Aufbau Virtueller Forschungsumgebungen wird in Deutschland zum einen von der Deutschen Forschungsgemeinschaft (DFG) und zum anderen vom Bundesministerium für Bildung und Forschung (BMBF) gefördert. Das BMBF förderte insbesondere die D-Grid-Initiative und das Projekt WissGrid. Auf europäischer Ebene erging 2009 im Rahmen des 7. Forschungsrahmenprogramms (FP7) ein Call zu „Virtual Research Communities“. Auch die Allianz der Wissenschaftsorganisationen und die Deutsche Initiative für Netzwerkinformation (DINI) beschäftigen sich thematisch mit VFU. Wie die Arbeitsgruppe der KZII (KZII 2011:

---

<sup>2</sup> <http://www.soeb.de/vfu-soeb-3/abschlusstagung/>, Zugriff am 07.10.2014.

B76ff.) feststellt, fehlt jedoch ein systematischer Überblick über den Stand der Entwicklung. Weder habe sich eine beherrschende Technologie etabliert, noch gebe es Referenzarchitekturen.

Für das Verbundprojekt „Virtuelle Forschungsumgebung für die Sozioökonomische Berichterstattung“ (*VFU soeb 3*) waren zum einen die Ergebnisse des unter I.2 genannten Teilprojekts „Kollaborative Datenauswertung und Virtuelle Arbeitsumgebung“ maßgebliche Voraussetzung für die Entwicklung der VFU. Zum anderen veranlasste das SOFI in 2010 eine technische Evaluation verfügbarer Grid-Technologie zur Umsetzung einer virtuellen sozialwissenschaftlichen Forschungsumgebung. An der Expertise arbeiteten Harry Enke (Astrophysikalisches Institut Potsdam), Patrick Harms (Abteilung Forschung und Entwicklung der Niedersächsischen Staats- und Universitätsbibliothek Göttingen) und Frank Dickmann (Abteilung Medizinische Informatik der Universitätsmedizin Göttingen). Ihr Bericht (Dickmann u.a. 2010), behandelt Anforderungen an eine virtuelle Arbeitsumgebung, deren Komponenten und technische Umsetzungsmöglichkeiten mit Hilfe der D-Grid-Infrastruktur laufen sollen. Diese Expertise zur Struktur und zum Aufbau einer virtuellen Forschungsumgebung wurde bei der Entwicklung der *VFU soeb 3* verwendet.

Beim Arbeitspaket 5 „Portal und Hosting“ und Arbeitspaket „Middleware“ konnte das Projekt nicht von Standard-Technologien für die Implementierung virtueller Forschungsumgebungen ausgehen, da sich dieser Bereich der E-Sciences noch stark im Fluss befindet. Es wurde daher bereits in der Planungsphase beschlossen, als Technologiebasis das Open-Source-Framework Liferay<sup>3</sup> einzusetzen, das einen Quasi-Standard im verwandten Bereich der Nutzerportale setzt. Liferay bietet auf der einen Seite die auch in VFU benötigte Kollaborations-Funktionalität (Wiki, Forum, etc.) und auf der anderen Seite ein Framework für die modulare Ergänzung der in VFU benötigten Funktionen zur generischen und zur Community-spezifischen Unterstützung des Forschungsprozesses.

Für den als besonders wichtig identifizierten Bereich des Forschungs-Datenmanagements wurde im Projekt zunächst eine existierende Lösung auf Basis von iRODS und Fedora<sup>4</sup> evaluiert. Sie erwies sich aber als ungeeignet, so dass stattdessen die bei der GWDG in Entwicklung befindliche modulare Datenmanagement-Lösung CDSTAR (Schmitt u.a. 2014) weiter entwickelt und an die Anforderungen der *VFU soeb 3* angepasst wurde.

Bei der Entwicklung der Metadaten schemata Studienbeschreibung und Datennutzung (AP 7) wurde auf die Elemente des Metadatenstandards DDI 2 (DDI Alliance 2014)<sup>56</sup> zurückgegrif-

---

<sup>3</sup> <http://www.liferay.com/de>, Zugriff am 07.10.2014.

<sup>4</sup> <https://wiki.irods.org/index.php/Fedora>, Zugriff am 07.10.2014.

<sup>5</sup> Data Documentation Initiative <http://www.ddialliance.org/>, Stand: 07.10.2014.

fen. Aktuell existiert jedoch noch keine systematisch zusammenhängende Metadatenstruktur zur ausführlichen Dokumentation von Syntaxdateien und Outputdateien und des projektbezogenen Kontextes von kollaborativ durchgeführten Datenanalysen (vgl. Friedhoff u.a. 2013: 12ff.). Aus diesem Grund wurde für die Bedürfnisse der *VFU soeb 3* das Metadatenschema „Syntax“ neu entwickelt.

Die Editoren (AP 8) wurden als Spring MVC Portlets (Version 3.2.0) für die Kooperationsplattform Liferay (Version 6.2) entwickelt. Ein Portlet ist ein Teilmodul innerhalb von Liferay und greift regelmäßig auf verschiedene Liferay-Funktionalitäten und andere maßgeschneiderte VFU-Module zu. Insbesondere gilt dies für die Datenhaltung und die Suche, deren Zusammenspiel im Folgenden erläutert werden soll. Die Module selbst wurden von der GWDG entwickelt, die für die Middleware-Entwicklung sowie die Basisfunktionalität des eingesetzten Portals zuständig war (vgl. Pohl 2014). Im Editor erstellte oder veränderte Objekte werden vom Editor in das JSON-Format gewandelt und an die Storage API geschickt. Nach Bedarf werden weitere Objekte aus dem Storage ebenso im JSON-Format gelesen und angezeigt und können dann im Editor bearbeitet werden. Sämtliche Operationen zum Lesen, Schreiben, Löschen und Verwalten von Inhalten laufen über die Storage-API, die Metadaten, Zugriffsberechtigungen und Bitstreams in einer NoSQL Datenbank speichert. Komplexere Datenstrukturen, wie Listen, Mengen und Unterobjekte werden durch Referenzierungen dargestellt.

### **I.5 Zusammenarbeit**

Die verbundinterne Zusammenarbeit wurde in einer Kooperationsvereinbarung geregelt, die die Zuwendungsempfänger untereinander im Juli 2012 geschlossen haben.

In der Projektlaufzeit gab es einen fachlichen Austausch mit dem Teilprojekt Informations- und Dateninfrastruktur (INF) des Sonderforschungsbereichs 882 – Von Heterogenitäten zu Ungleichheiten – an der Universität Bielefeld. Die dort vorgesehene Virtuelle Forschungsumgebung umfasst eine konventionelle Arbeitsplattform zur Datenverwaltung, eine Forschungsdatenplattform zur Archivierung, Nachnutzung und Dokumentation der Daten aller Teilprojekte und ein Schnittstellenmodul zur Anbindung an Forschungsdatenzentren und zum Datenaustausch.<sup>7</sup>

Darüber hinaus gab es einen Austausch mit dem Rat für Sozial- und Wirtschaftsdaten (RatSWD). Das Projekt wurde u.a. im Oktober 2013 auf der 10. Sitzung des ständigen Aus-

---

<sup>6</sup> DDI Empfehlungen zu kontrollierten Vokabularen: <http://www.ddialliance.org/Specification/DDI-CV/>, Stand 05.10.2014.

<sup>7</sup> <http://www.sfb882.uni-bielefeld.de/de/inf/vfu>, Zugriff am 07.10.2014.

schusses FDI des RatSWD) vorgestellt. Aus dem RatSWD haben Stefan Bender und Ulrich Rendtel auf der Abschlusstagung des Projekts *VFU soeb 3* gesprochen.

Mit dem Forschungsdatenzentrum der Statistischen Landesämter (Geschäftsstelle, Nordrhein-Westfalen) gab es ebenfalls einen fachlichen Austausch, und das FDZ beteiligte sich an der Abschlusstagung.

Zudem wurde bei der Entwicklung der Middleware (AP 6) mit den Projekten DARIAH-DE (Forschungsdatenmanagement) und dem INF-Projekt des SFB 963 (Forschungsdaten in der Astrophysik) zusammengearbeitet.

## II. Eingehende Darstellung

### II.1 Erzielte Ergebnisse

#### *II.1.1 Arbeitspaket 1: „Projektkoordination, Projektmanagement“ (SOFI)*

In Arbeitspaket 1 wurden die Entwicklungsarbeit der technischen Partneereinrichtungen GWDG, GESIS und D-Grid-gGmbH, die Beteiligung der FDZ und die Definition der fachlichen Anforderungen sowie die Evaluation der VFU-Komponenten aus Anwendungsperspektive koordiniert. Die projektinterne Kooperation der beteiligten Einrichtungen und die Kommunikation mit den fachwissenschaftlichen Anwender/inne/n wurden moderiert und unterstützt. Zu den Aufgaben gehörte im Detail u.a. das Monitoring von Meilensteinen, die Entwicklung und Abstimmung einer gemeinsamen Terminologie und eines allen Projektbeteiligten gemeinsamen Aufgabenverständnisses, die Definition geeigneter projektinterner Kommunikationsformate, wie Verbundtreffen, Arbeitstreffen, technische Arbeitsgruppe, Nutzungsstudien, die Planung und Organisation der Veranstaltungen, wie Auftaktworkshop und Abschlusstagung, die Definition von Nutzungsrollen der fachwissenschaftlichen Anwender/innen für die Rechteverwaltung, die Definition von Archivfunktionen, die Priorisierung von Aufgaben und die Administration der VFU in Zusammenarbeit mit der GWDG. AP 1 fungierte außerdem als Schnittstelle zum Verbundvorhaben *soeb 3*. In Zusammenarbeit mit Dr. Tanja Schmidt (AP 2) wurden auf der sozialwissenschaftlichen Seite die fachwissenschaftlichen Anforderungen an eine VFU spezifiziert und Nutzungsstudien für die entwickelten Komponenten der VFU konzipiert und durchgeführt.

Die Gesamtkoordination des Projekts lag beim SOFI: Dr. Peter Bartelheimer als Projektleiter und Sarah Cronjäger als wissenschaftliche Mitarbeiterin.

## II.1.2 Arbeitspaket 2: Fachwissenschaftliche Anforderungen, Anwendungsfälle (SOFI)

In Arbeitspaket 2 wurden typische fachliche Anwendungsfälle bei der kollaborativen Nutzung von Forschungsdaten aus der Praxis des Forschungsverbunds *soeb 3* heraus definiert und im Rahmen von Anforderungsanalysen in einzelnen Arbeitsschritten beschrieben. Die Spezifikationen wurden mit den Projektbeteiligten in den technischen Arbeitspaketen abgestimmt. Außerdem wurden die Anwendungsfälle, die in der technischen Expertise (Diekmann u.a. 2010) und im Abschlussbericht zur Konzeptstudie (Bartelheimer/Schmidt 2011) definiert wurden, weiter ausgearbeitet.

Dazu wurden auf Basis von Expert/innen-Interviews 27 typische Arbeitsprozesse als Use-Cases ermittelt und detailliert kategorisiert und beschrieben. Benötigte Funktionalitäten in der VFU wurden in Mock-Ups veranschaulicht. Abbildung 1 stellt beispielhaft das Mock-Up einer Ergebnisliste aus einer facettierten Suche dar.

Abbildung 1: Mockup Ergebnisliste aus facettierter Suche

Ergebnisliste - Archiv-Suche

Studie Datensatz Variablen Syntax Studienbeschreibung Datennutzung Person oder Arbeitspaket

Link zu Datennutzungsbeschreibung, Arbeitsdatensatz - Download - bei Berechtigung möglich  
Link zu Ergebnisliste der Person

Variablenname	Variablen- und Valuelabels	Arbeitsdatensatz	zugrundeliegende Syntax	Ausgangsvariable	AutorIn
<a href="#">..Bildgrupp</a>	<a href="#">Bildung gruppiert</a>	<a href="#">xpgen_bildtyp.dta</a>	<a href="#">bildung_recode_do</a>	<a href="#">bbpsbil</a>	<a href="#">Bartelheimer</a>
<a href="#">.Bildjahre</a>	<a href="#">Bildungsjahre</a>	<a href="#">aggregiertxp.sav</a>	<a href="#">aggregiertbildung.sps</a>	<a href="#">bbpsbil</a>	Schmidt
...					

link zur Stelle im do-file      link zum Metadatum: label/codebook der Variable      Link zu do-File im Syntax-Editor      link zu Metadaten der Originalvariablen

**Einschränkungen**

<input checked="" type="checkbox"/> Bartelheimer	<input checked="" type="checkbox"/> soep	<input checked="" type="checkbox"/> *.do
<input type="checkbox"/> Kohlrausch	<input type="checkbox"/> neps	<input checked="" type="checkbox"/> *.sps
<input checked="" type="checkbox"/> Schmidt	<input type="checkbox"/> pairfam	<input type="checkbox"/> *.R

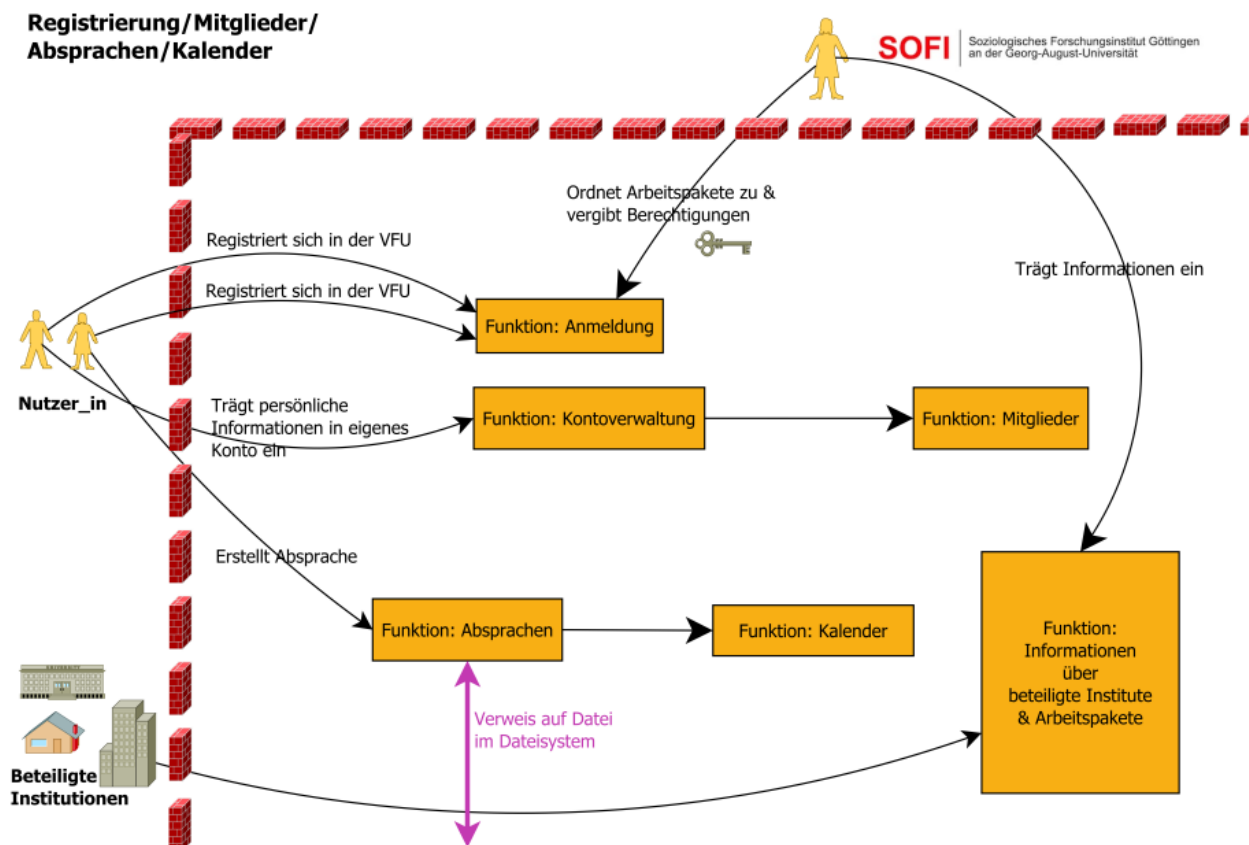
Die Use-Cases und die Arbeitsabläufe sowie die Mock-Ups wurden den technischen Arbeitspaketen übermittelt und im Rahmen verschiedener Arbeitstreffen vorgestellt und erläutert; auf dieser Grundlage wurden deren technische Umsetzungsmöglichkeiten diskutiert.

Zudem wurden umfangreiche Recherchen für benötigte Literaturverwaltungsprogramme durchgeführt und bereits verfügbare Open-Source-Produkte (beispielsweise RefBase<sup>8</sup>) dargestellt und dokumentiert.

Auf dem Auftaktworkshop des Forschungsverbunds soeb 3 im November 2013 wurden der aktuelle Entwicklungsstand der VFU (AP 5/6) und die geplanten Nutzungsstudien (AP 2) den Projektbeteiligten im Rahmen eines Workshops vorgestellt. Dazu wurden allen Teilnehmenden an verschiedenen "thematischen Tischen" die Archivfunktionen, die wichtigsten Portal-funktionen, das Dateiensystem sowie interne Verwaltungsoptionen live und interaktiv präsentiert.

Die folgenden vier Abbildungen (Abb. 2 bis 5) fassen den Entwicklungsstand der VFU zu diesem Zeitpunkt zusammen:

Abbildung 2: Registrierung, Mitglieder, Absprechen und Kalender der VFU



<sup>8</sup> [http://www.refbase.net/index.php/Feature\\_highlights](http://www.refbase.net/index.php/Feature_highlights), Stand: 06.10.2014.



Abbildung 3: Kollaboration im Wiki und Forum

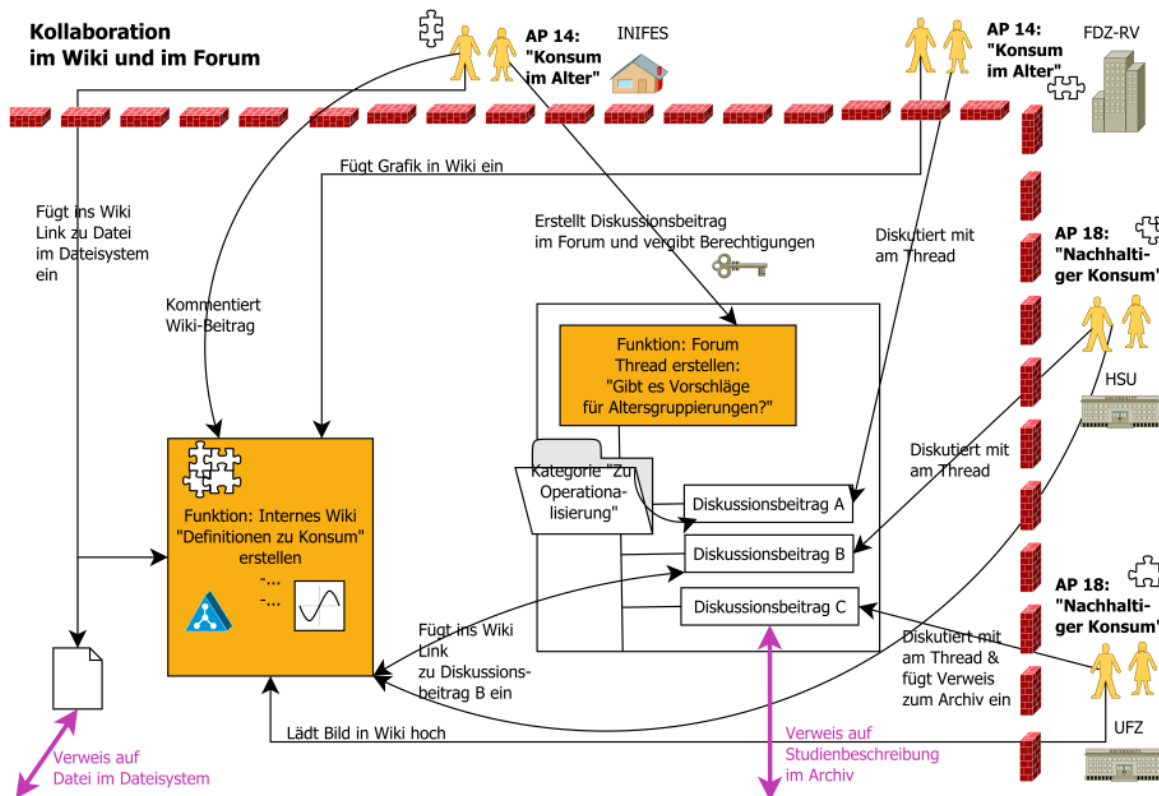


Abbildung 4: Dateisystem

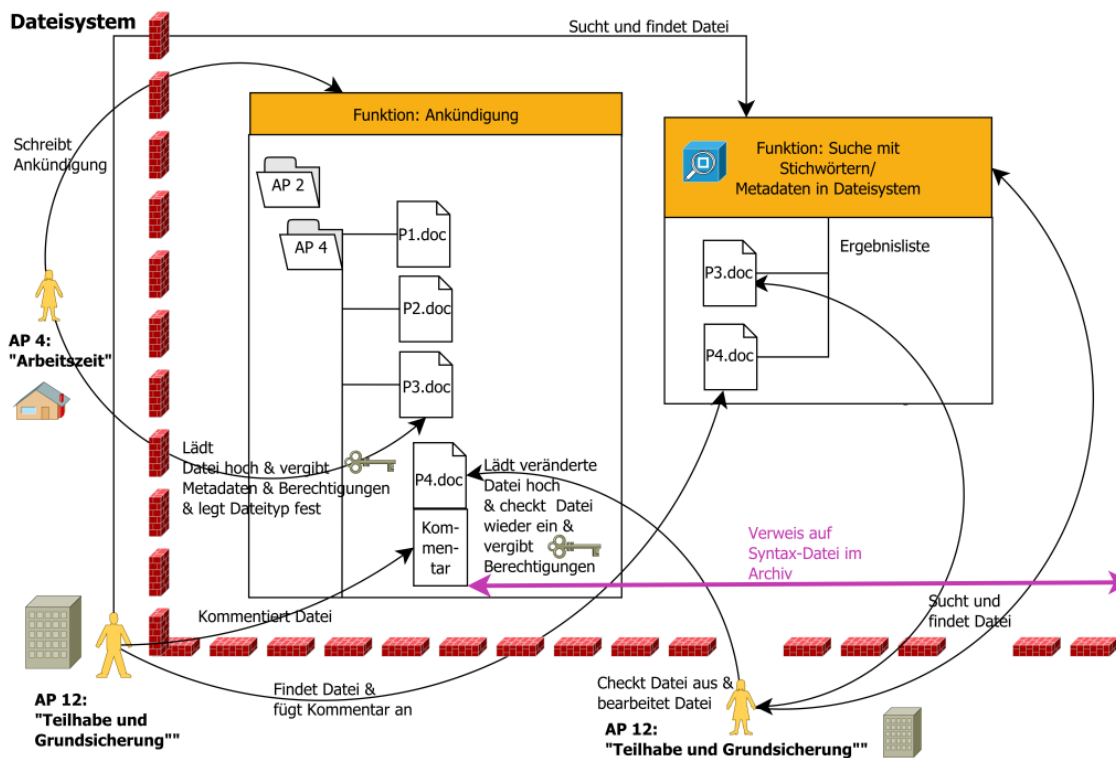
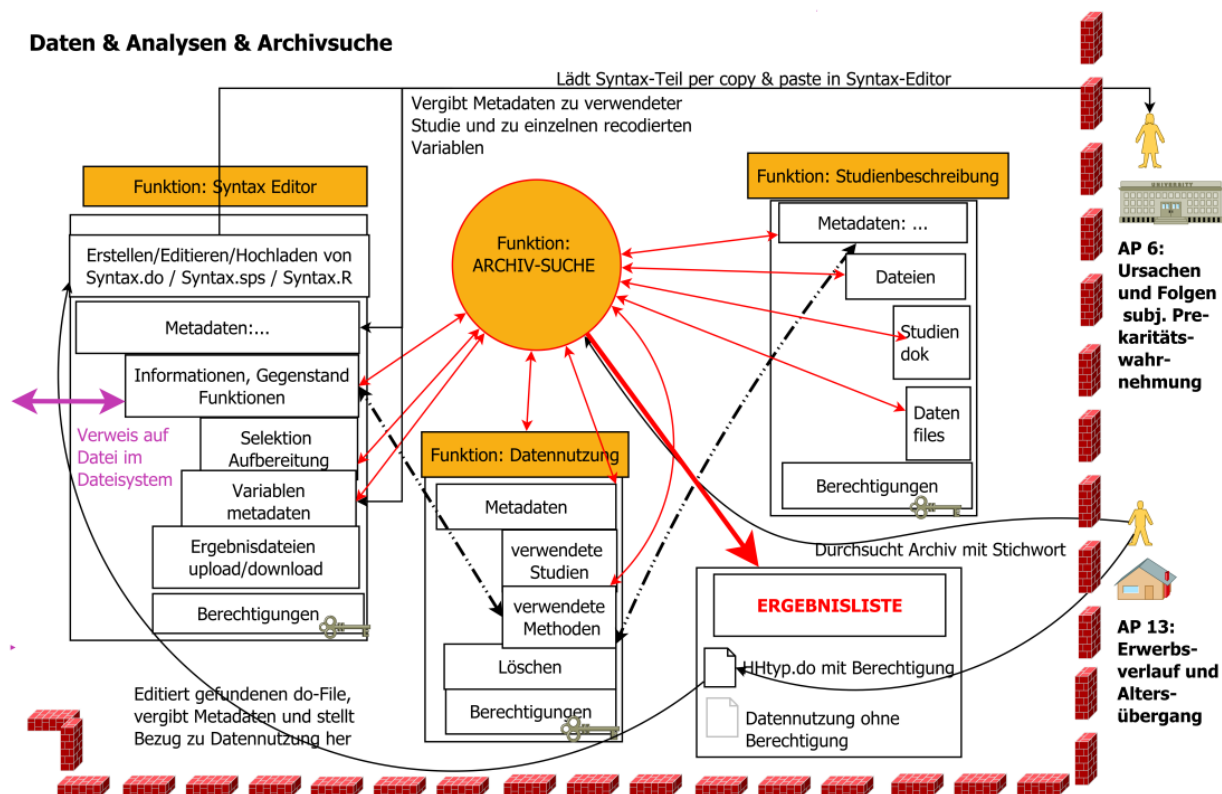


Abbildung 5: Daten, Analysen und Archivsuche



AP 2 hat die Gesamtarchitektur der VFU (AP 3), das Portal (AP 5), die Spezifikation von Metadatenstandards (AP 7) und die entwickelten Editoren (AP 8) fortlaufend aus Perspektive der fachwissenschaftlichen Nutzer/innen evaluiert. Darüber hinaus hat AP 2 zwei Nutzungsstudien für die im Projekt entwickelten Komponenten der VFU entwickelt und durchgeführt. In diesen Nutzungsstudien haben die Partnereinrichtungen des Forschungsverbunds soeb 3 über einen längeren Zeitraum unter Alltagsbedingungen Erfahrungen mit den Komponenten der VFU gemacht, die an die technischen Arbeitspakete des Infrastrukturprojekts rückgekoppelt wurden.

In der ersten Nutzungsstudie (Projektmonat 9 bis 12), die in Göttingen an einem Tag mit Testteilnehmer/innen durchgeführt wurde, wurden alle Funktionen der VFU als sinnvoll und nützlich beschrieben. Neben den positiven Aspekten der Funktionalitäten in der VFU wurden aber auch einige Verbesserungsmöglichkeiten für die VFU genannt. Dazu gehört vor allem der Navigation und Benutzerführung innerhalb verschiedener Funktionalitäten.

Im Rahmen der zweiten Nutzungsstudie (Projektmonat 15 bis 18) sollten die Teilnehmenden innerhalb von vier Wochen einige Aufgaben bearbeiten und dann online in einem Befragungstool bewerten. Während die quantitativen Bewertungen zur Navigation, Gestaltung und

Technik gut ausfielen, wurden gerade in den offenen Kommentaren Kritik und Unklarheiten bei dem Umgang mit verschiedenen Funktionen der VFU deutlich. Insbesondere im qualitativen Auswertungsteil wurden der hohe Zeitaufwand bei einzelnen Funktionen und umständliche Lösungen kritisiert. Damit waren neben den zeitintensiven Einzelclicks bei der Berechtigungsvergabe beispielsweise umständliche Verlinkungen gemeint, oder sich wiederholende Angaben zu den Metadaten im Bereich „Daten und Analysen“. Allerdings bekräftigten die positiven Anmerkungen, dass die Nutzer/innen sehr zufrieden mit dem Angebot an Funktionen sind. AP 2 hat beide Nutzungsstudien durchgeführt, ausgewertet und die Ergebnisse schriftlich in einem Paper dem Verbund vorgelegt. Darüber hinaus wurden die Ergebnisse im Verbund auf Arbeits- und Verbundtreffen präsentiert und diskutiert.

Die Vielzahl an Kommentaren, die Fragen, Missverständnisse und Unklarheiten der Befragten zum Ausdruck bringen, deutet insgesamt darauf hin, dass ein umfassendes Verständnis der VFU bei den Nutzer/innen noch Zeit braucht. Erst in der Zukunft, wenn ganze Arbeitsabläufe im Forschungsalltag beurteilen lassen, wird sich zeigen, ob und welcher zusätzliche Nutzen durch eine VFU für die Prozesse im Forschungsablauf erreicht werden kann.

AP 2 hat außerdem die Projektkoordination (AP 1) u.a. bei der Abstimmung mit dem Forschungsverbund *soeb 3*, bei der Erprobung der VFU und bei der Definition einer Archivstruktur unterstützt. Und schließlich wurde von AP 2 ein Benutzerhandbuch für die VFU erstellt, was dem Forschungsverbund vorgelegt wurde.

Das Arbeitspaket 2 wurde von Dr. Tanja Schmidt (Unterauftrag für SOFI), bei gleichzeitiger fachwissenschaftlicher Mitarbeit im Forschungsverbund *soeb 3*, bearbeitet.

### *II.1.3 Arbeitspaket 3: „Gesamtarchitektur“ (SUB)*

Aufgabe des Arbeitspaketes 3 war die Koordinierung der Gesamtarchitektur für den Aufbau der VFU *soeb 3* sowie die Unterstützung der Projektkoordination bei technischen Fragen. Hierin eingeschlossen waren die Entwicklung eines Konzepts zur Gesamtarchitektur, die Koordination der Arbeit der technischen Partner sowie die Entwicklung eines Konzepts zur Langzeitarchivierung.

AP 3 hat zu den fachwissenschaftlichen Anforderungen eng mit AP 1 und 2 und den beteiligten FDZ (AP 4) zusammen gearbeitet, sowie im technischen Bereich mit GWDG bei der Entwicklung von Portal (AP 5) und Middleware (AP 6), und mit GESIS bei den Metadaten (AP 7) und fachwissenschaftlichen Werkzeugen (AP 8). Darüber hinaus hat es die Projektkoordination (AP 1) bei Entscheidungsprozessen im technologischen Bereich, bei Priorisierungen sowie in Fragen der (technischen) Nachhaltigkeit und des künftigen Betriebs beraten. AP 1 und AP 2 wurden außerdem bei der Vermittlung und „Übersetzung“ von umgangs-

sprachlich oder fachwissenschaftlich formulierten Anforderungen in informationstechnische Konzepte unterstützt.

In die Entwicklung der Gesamtarchitektur für den Aufbau der *VFU soeb 3* flossen neben der VirtAug-Expertise (Dickmann u.a. 2010) auch die Ergebnisse aus Diskussionen mit den Projektpartnern auf der Auftaktveranstaltung sowie vorausgegangene Erfahrungen von SUB und D-Grid mit ein. Die FDZ wurden im Architekturschema zunächst über allgemeine Speicher- und Zugriffsschnittstellen an die VFU angebunden. Die weitere Entwicklung berücksichtigte hiernach die vom Partner GWDG vorgestellten und umgesetzten Lösungen zu Portal (Portal-Software „Liferay“) und Middleware (GWDG-eigene Archiv-Lösung „CD-Star“), sowie deren Möglichkeiten zur Anbindung der von den fachlichen Partnern gewünschten Funktionalitäten.

Technisch ergab sich hierbei insbesondere die Notwendigkeit für zwei getrennte Dateiverwaltungen von Portal und Archiv, was jedoch für die Gesamtarchitektur keine größeren Nachteile implizierte. Auf der Basis des entwickelten Architektur-Schemas konnten zur Veranschaulichung und zur internen Diskussion sowohl die von AP 2 entwickelten Anwendungsfälle abgebildet als auch Unklarheiten und potentielle Konflikte im Workflow und im Rechtemanagement identifiziert werden.

Die Moderation zwischen fachlichen und technischen Partnern erfolgte im Wesentlichen über die von AP 1 organisierten persönlichen Treffen. Hierbei wurden entweder konkrete größere Elemente der VFU diskutiert (z.B. Archiv-Suche, Editoren-Interface) oder die Priorisierung der anstehenden Implementierungsaufgaben diskutiert und vorgenommen. Hilfreich erwies sich hier die von AP 5 für die Entwicklungsarbeit zur Verfügung gestellte und technisch moderierte interne VFU-Instanz. Hier konnten in Forum und Wiki Fragen und Probleme diskutiert und nachverfolgt werden, sowie in der Dateiverwaltung für alle zugängliche Dateien strukturiert und auffindbar abgelegt werden. Im weiteren Projektverlauf wurde diese Struktur auf Vorschlag von AP 1 und AP 3 noch ergänzt um die von AP 8 genutzte Issue-Tracking Software „Mantis“, in welcher eine noch effizientere Kategorisierung, Zuordnung und Nachverfolgung der anstehenden Fehlerbehebungen und Feature-Wünsche möglich war.

Über die persönlichen Treffen und die Online-Kommunikation hinaus war die zweiwöchentliche, über die gesamte Projektlaufzeit hindurch von AP 3 organisierte Telefonkonferenz der technischen Arbeitsgruppe („TAG-Telko“) entscheidend für die durchgängige Abstimmung innerhalb der technischen APs in Bezug auf Ressourcen, Zeitpläne und Implementierungspläne. In Vorbereitung auf die Treffen wurden von jedem beteiligten AP aktuelle, kurze Statusberichte zu den erledigten und anstehenden Aufgaben erstellt und in das Wiki im internen VFU-Portal hochgeladen. Diese Berichte sowie eventuelle weitere Fragen wurden dann auf der TAG-Telko diskutiert. Anschließend wurden über die Berichte hinausgehende Absprachen und Ergebnisse in einem Protokoll festgehalten, welches ebenfalls im Wiki veröffent-

licht wurde und somit auch den anderen APs zur Verfügung stand. Zu den technischen AP 5 bis 8 trug AP 3 auch durch Tests der implementierten Komponenten bei. AP 3 erreichte somit eine koordinierte und nachvollziehbare Organisation des Implementierungsprozesses sowie die Abbildung der verfolgten Strategie in einer Gesamtarchitektur.

Das von AP 3 zum Projektabschluss erarbeitete Konzept zur Langzeitarchivierung beschreibt die erforderlichen Maßnahmen und Verantwortlichkeiten zur Langzeitarchivierung von Forschungsdaten und Ergebnissen in der *VFU soeb 3*. Es stellt hierbei zunächst die für die VFU relevanten Ziele von Langzeitarchivierung heraus und leitet daraus Anforderungen und zu realisierende Aufgaben ab. Abschließend geht es auf den aktuellen Stand der Archivierung in der *VFU soeb 3* ein und stellt die bereits vorhandenen Funktionalitäten den weiter bestehenden Anforderungen gegenüber.

Für das Arbeitspaket 3 war die SUB (Dr. Heike Neuroth, Timo Gnad, Jens Ludwig), zusammen mit der D-Grid GmbH (Prof. Dr. Uwe Schwiegelshohn) verantwortlich.

#### *II.1.4 Arbeitspaket 4: „Forschungsdaten“ (FDZ IAB, FDZ SOEP, FDZ RV)*

Die beteiligten Forschungsdatenzentren (FDZ IAB, FDZ RV, FDZ SOEP) haben in Arbeitspaket 4 ihre Daten- und Metadatenkompetenz, ihr Wissen aus den Bereichen Langzeitarchivierung, die organisatorischen und technischen Voraussetzungen des Zugangs zu ihren Mikrodaten (Remote Access und/oder Job Submission) sowie ihre datenschutzrechtliche Anforderungen für den Umgang mit Mikrodaten und den Ablauf wissenschaftlichen Arbeitens mit sensiblen Mikrodaten in die Entwicklung der VFU eingebracht. Sie haben die Projektkoordination und die technischen Arbeitspakete bei der Entwicklung einer Archivstruktur, eines Syntax-, Metadaten- und Studieneditors beraten.

Die für AP 4 im Projekt gesetzten Meilensteine konnten realisiert werden. Im Detail handelt es sich dabei um:

- die Definition von Schnittstellen für die Übermittlung von Programmcode (Job-Submission-Applikation),
- die Beratung des Gesamtprojekts zur Verwendung von Datensätzen der FDZ,
- die Bereitstellung von Metadaten zu diesen Datensätzen,
- die Mitarbeit am Datensicherheitskonzept und der Nutzungsordnung,
- die Unterstützung des Gesamtprojekts bei der Erstellung eines Metadatenstandards für die Verwendung in der VFU,
- Teilnahme an beiden Nutzungsstudien inkl. Feedbackrunden,
- Teilnahme an der Abschlussveranstaltung und Präsentationen,

- Beteiligung an Vorarbeiten zum Mikrodatenzugang aus der VFU (Proof-of-Concept-Präsentation auf der Abschlussveranstaltung).

Neben den realisierten Meilensteinen erfolgte auf verschiedene Ebenen eine kontinuierliche Beratungs- und Unterstützungstätigkeiten; im täglichen Betrieb genauso wie auf diversen Projekttreffen.

AP 4 hatte auch die Zugangswege zu Mikrodaten für eine VFU zu spezifizieren. Der Zugang zu besonders schützenswerten Mikrodaten kann aus datenschutzrechtlichen Gründen nur auf drei Arten gewährt werden: durch die Nutzung eines Gastwissenschaftler-Arbeitsplatzes in den Räumen der Dateneinrichtung, durch die Übermittlung von Programmcodes an die Dateneinrichtung durch Wissenschaftler/innen (Job Submission) und durch die Nutzung von sicheren Remote-Access-Verfahren. Für die Implementierung in die Struktur einer VFU kommen nur die beiden letzten Varianten in Frage.

In einem Treffen und in mehreren nachfolgenden Gesprächen wurden zwischen den FDZs und der GWDG Wege spezifiziert, um einen Programmcodes, der in der VFU entwickelt wurde, direkt an die FDZs senden zu können (Job Submission) und Ergebnisse der Berechnungen wieder in der VFU verfügbar machen zu können. Eine solche Lösung ist nicht nur auf ihre technische und organisatorische Machbarkeit, sondern auch auf ihre datenschutzrechtliche Anwendbarkeit hin zu prüfen. Der Datenzugangsweg Job Submission bietet durch seinen hohen Datenschutzstandard (Forscher/innen sehen die Mikrodaten zu keinem Zeitpunkt und erhalten nur datenschutzrechtlich geprüfte Ergebnisse zurück) die Möglichkeit der sicheren Analyse von Mikrodaten. Eine große Zahl von Datenbeständen kann so verfügbar gemacht werden. Leider birgt der Ansatz auch bedeutende Nachteile in sich. Dass die nutzenden Wissenschaftler/innen nie die Forschungsdaten sehen, macht effektive Forschungsarbeit schwer. Die Folge ist eine Vielzahl von eingesendeten Programmcodes, die immer eine Datenschutzprüfung notwendig machen und somit einen hohen Arbeitsaufwand auch auf der Seite der Datenanbieter erzeugen. Im Gegensatz zu typischen Job-Submission-Lösungen bietet die VFU hier den Vorteil einer geschützten Umgebung für die Betrachtung der Ergebnisse. Wie sich dieser Umstand auf die Datenschutzprüfung auswirkt, muss noch genauer geklärt werden.

Ein für die Forschung komfortablerer Weg der Datenanalyse ist der Zugang durch sicheren Remote Access (auch als Remote Desktop bezeichnet). Hierbei arbeiten die Nutzer/innen direkt mit den auf den geschützten Servern der Datenhalter gespeicherten Daten. Sie können somit „live“ den Datensatz erforschen und Zwischenergebnisse erzeugen. Dabei wird den Forschenden immer nur ein Bild (livestream) ihrer Arbeit von dem entfernten Server übermittelt. Nie wird der gesamte Datenbestand angezeigt oder gar übermittelt. Für die Wissenschaftler/innen handelt es sich hierbei um den Datenzugangsweg, der ihrer gewohnten

Arbeitsweise nahezu entspricht. Aufbauend auf der Remote-Access-Lösung des FDZ-IAB („FDZ-in FDZ-Ansatz“) wird an einer Weiterentwicklung im Rahmen der Implementierung eines Remote Access Zugangs in die VFU gearbeitet.

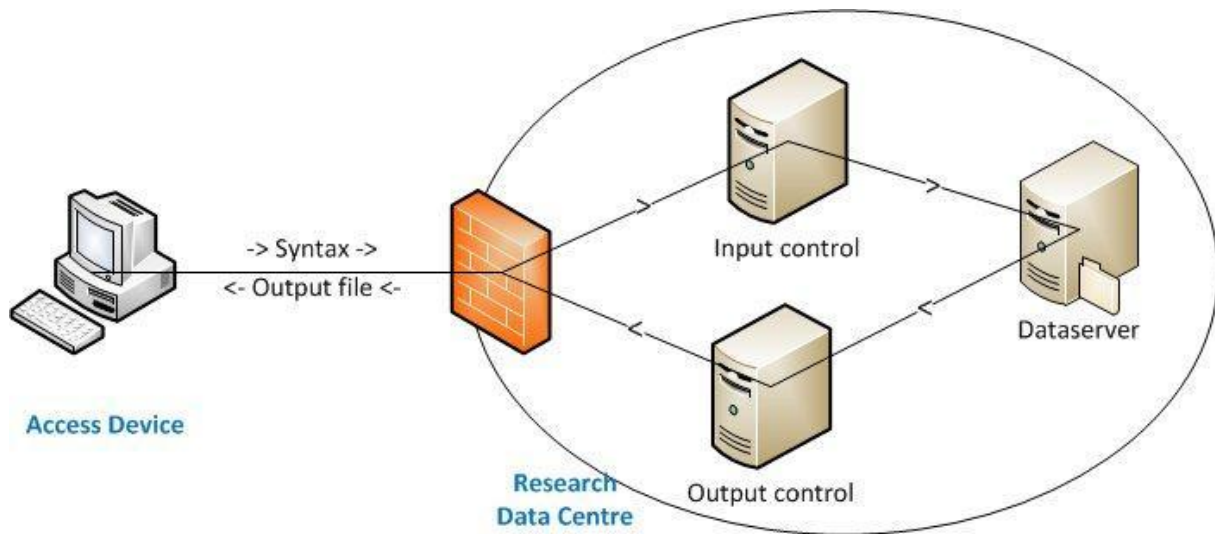
Im Bereich Metadaten waren die Forschungsdatenzentren vor allem an der Dokumentation von Programmcodes („Syntax“) für die vom Syntaxeditor unterstützten Statistikprogramme beteiligt. Für die Dokumentation von Forschungsdaten besteht bereits eine Vielzahl von Standards. Das Ziel ist hier eine Harmonisierung diese Standards. Für eine standardisierte Dokumentation des Programcodes, mit dem Forschungsdaten analysiert werden, konnte nicht an einen entwickelten Arbeitsstand in den Sozial- und Wirtschaftswissenschaften angeknüpft werden. Hier wurde im Rahmen des VFU-Projekts bedeutende Grundlagenarbeit geleistet.

Im Verlauf des Projekts trat die Einbindung von Forschungsdaten für AP 4 in den Vordergrund. Die Möglichkeiten des Datenzugangs via Job Submission wurden weiter erörtert. Eine schematische Darstellung dieses Zugangsweges gibt Abbildung 6. Entscheidend für die Einbindung in die VFU sind hier die Schnittstellen zu den geschützten Bereichen der beteiligten Forschungsdatenzentren. Wenn auch im Grunde stets der gleiche organisatorische Ablauf durchgeführt wird, so unterscheiden sich die Lösungen doch durch ihre spezifische technische Umsetzung. Da bei Job Submission nie die eigentlichen Mikrodaten betrachtet werden, sondern lediglich Ausgabedateien mit Ergebnissen statistischer Anfragen übermittelt werden, ist eine Einbindung in die VFU vom Grundsatz her leichter möglich. Die Anpassung an die verschiedenen Schnittstellen der unterstützten Forschungsdatenzentren ist aber zeit- und arbeitsaufwändig. Aufgrund dessen wurde entschieden, Job Submission als wichtigen und gangbaren Weg des Datenzugangs zu dokumentieren, die weitere praktische Entwicklungsarbeit im Projekt auf den Datenzugang via Remote Desktop auszurichten.

Remote Desktop stellt für die Daten nutzenden Wissenschaftler/innen den angenehmeren Weg des Datenzugangs dar, da er eher der üblichen Arbeitsweise mit Daten auf lokalen Rechnern entspricht. Eine schematische Darstellung dieses Zugangsweges zeigt Abbildung 7. Die Mikrodaten verbleiben auf den Servern der Dateneinrichtung; über eine verschlüsselte Internetverbindung werden lediglich Anfragen (via Tastatur und Maus) und graphische Rückmeldungen (via Bildschirmupdates) übermittelt. Die Nutzenden können die Daten aber live betrachten und bearbeiten, etwa Berechnungen anstellen.

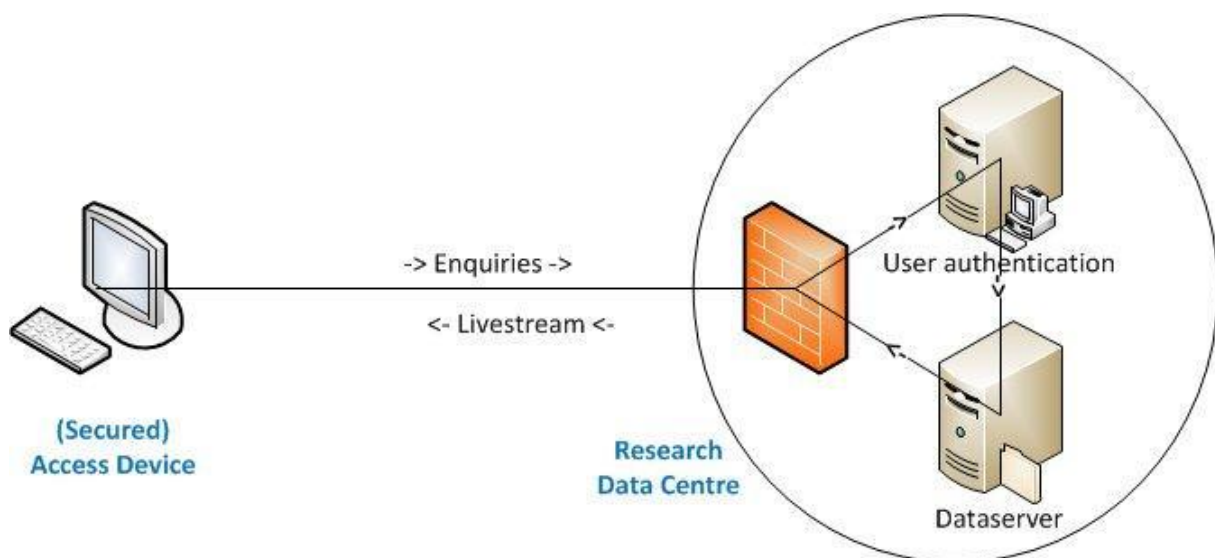
Auch die ursprünglich geplante Einbindung der quelloffenen Software R in die VFU wird durch Remote Desktop entbehrlich, da gängige Statistiksoftware auf den Rechnern des jeweiligen FDZ bereit gestellt wird.

Abbildung 6: Job Submission



In einer möglichen weiteren Entwicklungsphase soll im Rahmen einer zu realisierenden Remote-Data-Access-Funktionalität eine interaktive Nutzung von Statistikprogrammen auf GWDG-Ressourcen (Statistik-Arbeitsplatz) ermöglicht werden. Die Frage, ob eine Einbindung von Remote-Data-Access in die VFU überhaupt möglich ist, war bei diesem Zugangsweg und im Gegensatz zu Job Submission ungeklärt. Daher fiel die Entscheidung, im Rahmen eines Proof of Concept diesen Zugangsweg in die VFU einzubinden. Dieses Proof of Concept (eine technische Machbarkeitsstudie in einer geschützten IT-Umgebung) konnte erfolgreich abgeschlossen werden, und die Ergebnisse wurden auf der Abschlussveranstaltung des Projekts vorgestellt.

Abbildung 7: Remote Desktop





Die Entwicklung des Proof of Concept geschah zentral durch die GWDG und das IAB. Die anderen beteiligten FDZs lieferten wichtigen Input und Feedback. Die Einbindung des Zugangs zu Forschungsdaten (Job Submission wie Remote Desktop) in die VFU-Umgebung sollen begleitend in einer eigenen Publikation dargestellt werden<sup>9</sup>.

In der Projektlaufzeit wurden wichtige Dokumentationen und Pilotlösungen geschaffen. Eine echte Einbindung des Datenzugangs zu Forschungsdatenzentren in Deutschland bedarf weiterer Ressourcen für die technische Entwicklung und für die datenschutzrechtliche Prüfung. Ohne Einbindung des Zugangs zu Forschungsdaten bleibt die Unterstützung des Forschungsprozesses der quantitativ-empirisch arbeitenden Sozial- und Wirtschaftswissenschaften jedoch unvollständig. Mit einer solchen Einbindung dagegen können viele Infrastrukturkosten eingespart werden, da der jeweilige Service nun zentral über die VFU angeboten werden kann.

Das Arbeitspaket 4 wurde vom FDZ der BA im IAB (Stefan Bender, David Schiller), dem FDZ-SOEP (Dr. Jan Goebel) und dem FDZ-RV (Tatjana Mika, Janine Lange) bearbeitet.

#### *II.1.5 Arbeitspaket 5: „Portal und Hosting“ (GWDG)*

Arbeitspaket 5 sollte eine Basis-VFU mit den wichtigsten kollaborativen Werkzeugen entwickeln. Die mit Liferay mitgelieferten Module (im Bereich der Nutzerportale Portlets genannt) für Foren und Wikis konnten für die VFU übernommen werden. Notwendige Anpassungen betrafen vor allem die Formulierungen in der Nutzerschnittstelle, deren Implementierung sich auch technisch als aufwändig erwies, da Liferay hierzu kein einheitliches Verfahren bietet. Ein Modul für die Absprache von Aufgaben in Projekten musste komplett neu entwickelt werden, da der Funktionsumfang vorhandener Lösungen unzureichend war.

Die mit Liferay mitgelieferte Dateiverwaltung wurde als komplementäre Lösung zum neu implementierten Forschungsdatenmanagement ebenfalls übernommen. Sie bietet nicht die aufgrund der Anforderungen von AP 7 und 8 erforderliche Funktionalität bezüglich der Behandlung von Metadaten, ist aber gut für das Management von Arbeitsdokumenten einschließlich deren einfacher Verlinkung in den anderen Portlets geeignet. Auch hier waren Anpassungen der Konfiguration und der Nutzerschnittstelle erforderlich.

Ein weiteres Ziel von Arbeitspaket 5 war die Implementierung von Authentifizierung und rollenbasierter Autorisierung. Für die RBAC (Role Based Access Control) in der VFU konnte die entsprechende Funktionalität aus Liferay verwendet werden, wofür die Enterprise-Edition von

---

<sup>9</sup> Working-Paper beim RatSWD; ist in Arbeit.

Liferay benötigt wurde. Die Implementierung erforderte die Konfiguration und Anpassung der LDAP-konformen Bereitstellung der erfassten Informationen zu Nutzern und ihren Rollen für die Middleware CDSTAR. Weitere durchgeführte Arbeiten betrafen vor allem die Umsetzung des durch die Projektleitung vorgegebenen Rollenkonzeptes in der Liferay RBAC.

Eine weitere Aufgabe von AP 5 war die Integration der in AP 8 entwickelten Module in Liferay. AP 5 übernahm den Support des Projektpartners GESIS bei der Integration seiner Module in die von der GWDG betriebene Liferay-Infrastruktur, einschließlich Bereitstellung erforderlicher Zugänge und den Anpassungen an den Test- und Produktivumgebungen, sowie Unterstützung bei den Funktionstests. Notwendige Änderungen am Datenmanagement werden unter AP 6 erfasst.

In AP 5 wurde außerdem eine Hosting-Infrastruktur aufgebaut und gehostet. Es wurden eine Produktiv-, eine Staging- (Test-) und eine Entwicklungsinstanz der VFU auf virtuellen Maschinen eingerichtet und bereitgestellt. Hinzu kamen eine Test- und eine Entwicklungsinstanz für die CDSTAR-Middleware. Die Wartung der Systeme umfasste Pflege und Patching des Betriebssystems und der Basis-Komponenten der VFU (Apache Webserver, JBOSS Application Server, SQL Datenbank) sowie das notwendige Backup.

AP 5 war darüber hinaus an der Erstellung des Proof-of-Concept für den Remote Data Access in Zusammenarbeit mit AP 4 beteiligt (siehe Ergebnisse AP 4). Hierzu wurde auf Basis des Open-Source-Produktes Guacamole<sup>10</sup> ein funktionaler Prototyp ohne Integration in die VFU erstellt.

Schließlich hatte AP 5 das Update des Portals auf die aktuelle Liferay-Version 6.2 und das geplante Neudesign der optischen Elemente der Nutzerschnittstelle (Face-Lifting) vorzubereiten. Für das Upgrade der VFU auf die aktuelle Liferay Version 6.2 und ein damit einhergehendes neues Design wurden umfassende Tests mit den vorhandenen Modulen durchgeführt sowie notwendige Anpassungen implementiert. Die Arbeiten wurden im Rahmen der kostenneutralen Verlängerung des Projektes durchgeführt, konnten aber noch nicht abgeschlossen werden. Die endgültige Implementierung des Upgrades erfolgt auf Basis einer Eigenbeteiligung der GWDG nach Abschluss des Projektes.

Das Arbeitspaket 5 wurde von der GWDG (Prof. Dr. Ramin Yahyapour, Dr. Philipp Wieder, Dr. Christian Boehme, Christof Pohl) durchgeführt.

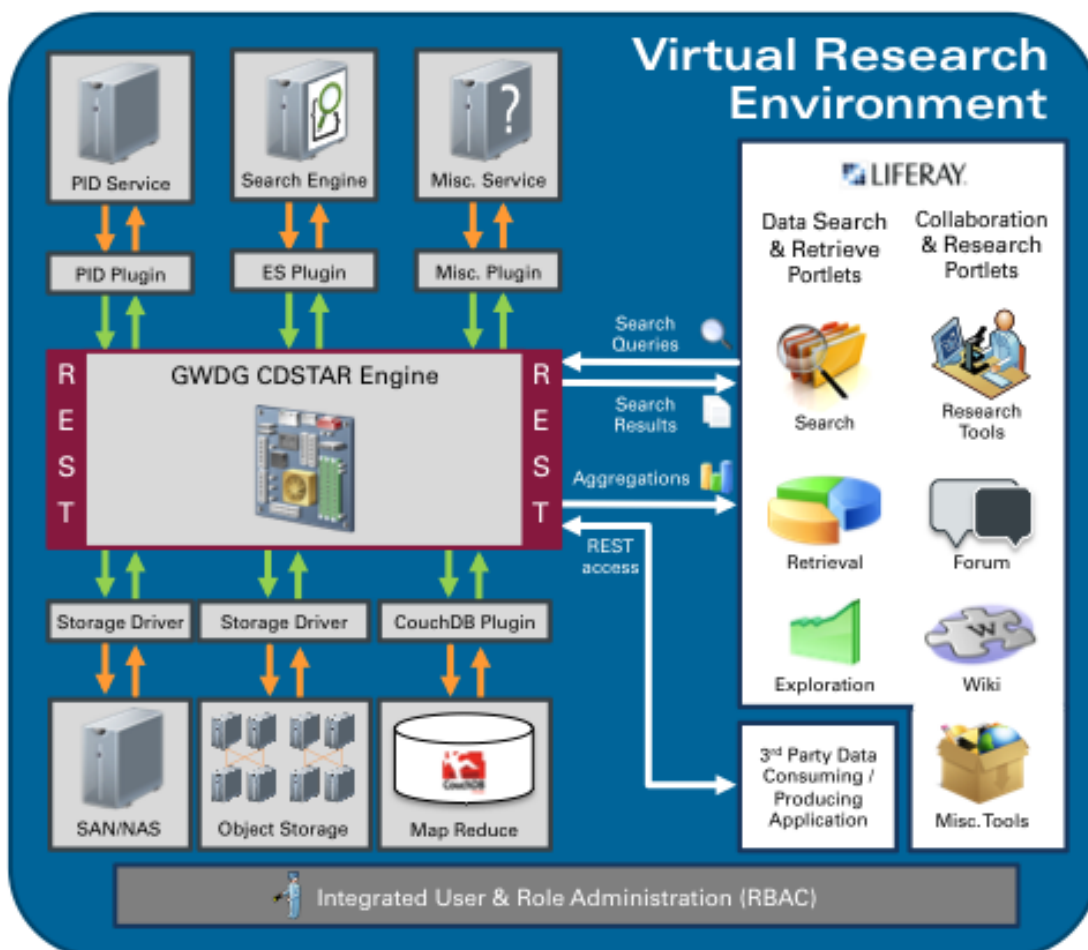
---

<sup>10</sup> <http://guac-dev.org>, Stand: 06.10.2014.

### II.1.6 Arbeitspaket 6: „Middleware“ (GWDG)

Arbeitspaket 6 kam die Aufgabe zu, ein Anforderungsprofil für das Datenmanagement zu erstellen und eine geeignete Middleware auszuwählen. Die als Middleware vorgesehene Kombination aus iRODS und Fedora erwies sich in der Evaluation als zu komplex, schwer zu warten, schwer anzupassen und als zu wenig stabil. Daher wurde dem parallel evaluierten, bei der GWDG in Entwicklung befindlichen, CDSTAR-Projekt der Vorzug gegeben. CDSTAR zeichnet sich vor allem durch seinen modularen Ansatz aus, der es auf einfache Weise ermöglicht, benötigte Features wie Volltext-Suche, Metadatenbank und zukünftig Unterstützung von Archivmöglichkeiten, Persistent Identifiers, und weitere Werkzeuge für das Datenmanagement zu integrieren. Das im Projekt umgesetzte Konzept zur modularen Integration von Liferay, CDSTAR und ihren jeweiligen Subkomponenten zeigt Abbildung 8.

Abbildung 8: Komponenten-Architektur der VFU soeb 3



Ein weiteres Ziel von AP 6 war die Konzeption der notwendigen Anpassungen von CDSTAR an die Anforderungen der VFU soeb 3 und die Spezifikation der Schnittstellen zum Portal. Die Definition der benötigten Schnittstellen zwischen Portal und Middleware musste insbe-

sondere die Integration des Identitätsmanagements (RBAC) und der gewünschten Funktionalität bezüglich der Erschließung von Metadaten berücksichtigen. Dies waren auch die entscheidenden Punkte im Konzept zur Weiterentwicklung von CDSTAR für die *VFU soeb 3*.

In AP 6 wurden außerdem Anpassungen an die in AP 8 entwickelten Anwendungen vorgenommen. Die Anforderungen der Anwendungslogik der in AP 8 entwickelten Module zum Management von Forschungsdatensätzen („Studien“) und Syntaxen und deren Metadaten wurden in der Basisversion nicht optimal unterstützt. Insbesondere waren manche Funktionen der Module nur umständlich auf CDSTAR abbildbar. Hier wurden eine Reihe von Optimierungen und Verbesserungen implementiert.

Ein weiteres Ziel von AP 6 war die Implementierung einer Suchfunktion. Die Anforderungen des Forschungsverbunds *soeb 3* erforderten Suchmöglichkeiten sowohl in den Daten als auch in den Metadaten von gespeicherten Datenobjekten. Als Basis dieser Funktionalität wurde die quelloffene Enterprise-Search-Engine ElasticSearch<sup>11</sup> verwendet. Für die Integration war vor allem die Implementierung einer Schnittstelle für die Suchmaschine in CDSTAR notwendig, des Weiteren die Konzeption und Implementierung einer geeigneten Konfiguration von ElasticSearch.

In AP 6 fand ebenfalls die Implementierung von RBAC und dem *soeb 3* Rollenkonzept statt. Die Ausgangsversion von CDSTAR enthielt nur Basisfunktionen für Authentifizierung und Autorisierung. Die komplexen Anforderungen eines RBAC machten erhebliche Änderungen an CDSTAR erforderlich, die es ermöglichen, RBAC Regularien auf allen Ebenen der VFU umzusetzen. Des Weiteren mussten die spezifischen Anforderungen des Rollenkonzeptes von *soeb 3* implementiert werden.

AP 6 setzte außerdem die Implementierung des Metadatenschemas aus AP 7 um, was zusätzliche Funktionalität in der Metadatenbehandlung in CDSTAR erforderte.

Das Hauptergebnis des Projektes aus technischer Sicht ist eine VFU, die aus einem modularen Portal (Liferay) (AP 5) und einer modularen Datenmanagement Middleware (CDSTAR) (AP 6) besteht. Dieser durchgängig modulare Ansatz erlaubt es, die vorhandene VFU als Basis für zukünftige VFU-Projekte zu verwenden. Hinsichtlich der Wiederverwendbarkeit ist zu unterscheiden zwischen generischen Modulen und fachspezifischen Modulen. Zu den generischen Modulen zählen die in Liferay integrierten Kollaborationstools, aber auch Funktionalität des Datenmanagements, wie zum Beispiel die Suchfunktion in Daten und Metadaten, die RBAC oder das Metadatenmanagement. Solche generischen Module können in zukünftigen VFUen weiterverwendet werden, wobei typischerweise anwendungsspezifische Anpassungen erforderlich sein dürften. Fachspezifische Funktionalitäten, wie beispielsweise

---

<sup>11</sup> <http://www.elasticsearch.org>, Stand: 06.10.2014.

die Editoren (AP 8), werden auch für andere Communities erforderlich sein, und werden im Regelfall neu entwickelt werden müssen.

Als technisches Nebenergebnis ist die erreichte hohe Reife der Middleware CDSTAR zu nennen. CDSTAR wird nun auch in anderen Projekten erfolgreich eingesetzt.

Das Arbeitspaket 6 wurde von der GWDG (Prof. Dr. Ramin Yahyapour, Dr. Philipp Wieder, Dr. Christian Boehme, Christof Pohl, Stephan Hilker, Oliver Schmitt) durchgeführt.

### *II.1.7 Arbeitspaket 7: „Metadaten“ (GESIS)*

Arbeitspaket 7 sollte die Metadatenelemente bestimmen, die in der *VFU soeb 3* benötigt werden. Dieses Ziel wurde erreicht, indem drei Metdatenschemata so spezifiziert wurden, dass die Anforderungen an die Funktionen einer VFU für kollaboratives Arbeiten mit diesen Metadaten erfüllt werden können.

Die Entwicklung der Metdatenschemata erfolgte in drei Stufen. Zunächst wurde ein grundlegendes Set entwickelt. Nach fachlicher Rückmeldung der Projektpartner und Nutzungsstudien wurde das grundlegende Metadatenset schrittweise zu einem erweiterten und schließlich zu einem finalen Metadatenset ausgearbeitet.

Es wurden drei technisch und inhaltlich spezifizierte Metdatenschemata entwickelt. Diese beziehen sich auf die Objekte Studie, Datennutzung und Syntax:

- a) Das in Archivkontexten gebräuchliche Metdatenschema „Studienbeschreibung“ dokumentiert die Datensätze von Datenanbietern, die für Datenanalysen im Verbundprojekt genutzt werden. Die Metadaten umfassen Angaben des Anbieters zu Bibliographie, Datensatz, Datenschutz, Nutzungsrechte, Methoden und Dokumente. Sie unterstützen die nachhaltige Sicherung von Herkunfts- und Kontextangaben über die im Verbundprojekt genutzten Forschungsdaten (z.B. Persistent Identifier bzw. Fingerprint des Datensatzes).
- b) Das Metdatenschema „Datennutzung“ erfasst u.a. Untersuchungsgegenstände und methodische Aspekte, die die Forscher/innen in den *soeb* Arbeitspakete den Datenanalysen zugrunde legen. Referenzierte Metadaten mit Angaben der einzelnen Arbeitspakete erleichtern das kollaborative Arbeiten z.B. in Recherchen, indem Untersuchungsgegenstände überschaubar abgebildet werden.
- c) Das Metdatenschema „Syntax“ dokumentiert Datenanalysen mit Statistikprogrammen durch Syntaxdateien und Ergebnisdateien, die von Mitarbeiter/inne/n eines Forschungs-Arbeitspaketes erstellt werden. Sie erleichtern das kollaborative Arbeiten, indem sie eine Übersicht über vorhandene Syntaxen und Entwicklungsstände innerhalb einzelner wissenschaftlicher Arbeitspakete geben. Weitere Metadaten (Freigabe;

Nutzungsbedingungen) ermöglichen das Teilen von Syntaxen über wissenschaftliche Arbeitspakete hinaus.

Die drei Schemata beschreiben ausschließlich Objekte und die mit ihnen verknüpften Dateien, die Gegenstand üblicher Arbeitsabläufe bei der Durchführung von Datenanalysen sind. Die Erfassung und Erschließung von analysebezogenen Metadaten in der *VFU soeb 3* unterstützt damit übergreifend die kollaborative Nutzung sozial- und wirtschaftswissenschaftlicher Mikrodaten.

Mit Hilfe dieser Metadaten können die Forschenden u.a. in Analysen genutzte Ausgangsdaten, Untersuchungsziele, methodische Strategien und Analyseparameter der Datennutzung, angewendete statistische Methoden und konkret benötigten Syntaxen und deren Entwicklung systematisch dokumentieren. Gleichzeitig fördern die Metadaten die informationelle Vernetzung zwischen den wissenschaftlichen Arbeitspaketen, da z.B. alle von den Arbeitspaketen dokumentierten Datensätze im VFU-Archiv nach Metadaten durchsucht werden können. Insbesondere fördern die Mehrwertdienste des Archivs auf Grundlage der erfassten Metadaten die gemeinsame Entwicklung, den Austausch und die Nachnutzung von Auswertungssyntaxen. Im Sinne der Sicherung guter wissenschaftlicher Praxis dokumentieren die Metadaten der *VFU soeb 3* relevante Information des Forschungsprozesses von den Ausgangsdaten bis zu den Ergebnissen der Datenanalysen. Diese Angaben in den Metadaten werden nachprüfbar, nachnutzbar und dauerhaft im VFU Archiv gesichert und stehen damit auch über das Projektende hinaus für weitere Infrastrukturzwecke zur Verfügung. Während für die Entwicklung der Schemata Studien und Datennutzung auf das Metadatenschema DDI 2 zurückgegriffen werden konnte, existiert für das Schema zu Syntax kein vergleichbarer Standard. Aus diesem Grund wurde das Schema Syntax neu entwickelt.

Das Arbeitspaket 7 wurde von der GESIS (Uwe Jensen, Stefan Schweers) bearbeitet.

#### *II.1.8 Arbeitspaket 8: „Kollaborative Werkzeuge“ (GESIS)*

Aufgabe von AP 8 war die Entwicklung der kollaborativen Werkzeuge Syntaxeditor und Metadateneditor. Zur Erreichung dieses Ziels wurden die Editoren gemäß des Metadaten-Schemas aus AP 7 programmiert, wobei der Metdateneditor in die Module Studien, Datennutzung und Datendateien unterteilt wurde. Neben der Erfassung von Metadaten wurde für Studien und Datendateien eine Upload-Funktionalität, beispielsweise für Dokumente oder Dateien zur Studie, implementiert, um auch weiterführende Materialien zu Studien und Datendateien in der VFU mit anderen Nutzer/inne/n teilen und nutzen zu können.

Die Entwicklung der Editoren erfolgte in enger Zusammenarbeit mit den fachlichen Arbeitspaketen. Um ein gemeinsames Verständnis über die Aufgaben der Editoren zu erhalten, wurde zur Anforderungsermittlung auf Hilfsmittel wie Use Cases und MockUps zurückgegrif-

fen. Nach Abstimmung der Anforderungen wurden diese in lauffähige Prototypen überführt und kontinuierlich verbessert. Es hat sich gezeigt, dass eine nutzergetriebene Entwicklung zu einem gemeinsamen Verständnis über die Anforderungen führt und den Entwicklungsprozess unterstützt. Weiterhin erwiesen sich die Nutzungsstudien als hilfreich, um weitere Anforderungen zu ermitteln. So zeigte das Feedback aus den Nutzungsstudien, dass eine erleichterte Nutzung der Editoren gewünscht ist. Dazu zählt eine Übersicht der zuletzt bearbeiteten Objekte, die Verwendung kontrollierter Vokabulare und Tooltips als Hilfsmittel für die Arbeit mit Metadaten. Wie bei der Anforderungsermittlung aus den fachlichen Arbeitspaketen wurde auch das Feedback der Nutzungsstudien in den weiteren Entwicklungsprozess mit einbezogen.

Der Syntaxeditor umfasst kollaborative Features wie die Bearbeitung, das Hochladen und die Versionierung von Syntax. Daneben wurden Features implementiert, die die Nutzungsfreundlichkeit des Syntaxeditors erhöhen, indem aus Offline-Editoren bekannte Funktionen wie Syntax-Highlighting, Suchen-Ersetzen oder ein Anpassen der Schriftgröße implementiert wurden. Weitere Features des Syntaxeditors sind der Upload von Ergebnisdateien und Log-Files sowie die Dokumentation selbstgenerierter Variablen.

Sowohl der Metadateneditor als auch der Syntaxeditor implementieren das Rechte- und Rollenmodell der *VFU soeb 3*, so dass Berechtigungen für die jeweiligen Objekte vergeben werden können. Hierbei kann gewählt werden, ob die Rechte einzelnen Personen, Arbeitspaketen oder Rolleninhabern zugewiesen werden sollen. Unterschieden wird zwischen Lese-, Schreib- und Verwalten-Rechten an den Objekten.

Die im Rahmen dieses Projekts entwickelten Editoren ermöglichen kollaboratives Arbeiten im Rahmen des *VFU-Prototyps*. Durch die Verwendung von JSON als Ablage-Struktur ist jedoch eine nachträgliche Anpassung des Metadatenschemas mit geringem Aufwand möglich. Auf diese Weise lässt sich das Datenmodell auch um künftige Anforderungen erweitern. Änderungen an dem Metadatenschema müssen in der Regel auch erfassbar sein. Dies müsste in den Editoren, sowohl im Frontend als auch im Backend, noch nachimplementiert werden.

Das Arbeitspaket 7 wurde von der GESIS (Dr. Brigitte Mathiak, Zeljko Carevic, Wolfgang Zenk-Möltgen) bearbeitet.

## **II.2 Zahlenmäßiger Nachweis**

[vertraulich]

### II.3 Notwendigkeit und Angemessenheit der geleisteten Arbeit

Die Arbeitsplanung im Projektantrag und die entsprechend kalkulierten Kosten der Aufwände der einzelnen Projektpartner erwiesen sich im Projektverlauf, gemessen an den Projektzielen und den dargelegten erzielten Projektergebnissen der einzelnen Projektpartner, als notwendig und angemessen. Die erzielten Projektergebnisse waren nur durch die interdisziplinäre Zusammensetzung des Verbunds und in arbeitsteiliger, koordinierter Aufgabenwahrnehmung durch die Verbundpartner möglich.

### II.4 Voraussichtlicher Nutzen

Dass die *VFU soeb 3* als operativ einsetzbarer Prototyp entwickelt wurde und bei den Pilotanwendern des Forschungsverbunds *soeb 3* hohe Akzeptanz findet, zeigt den wissenschaftlichen Erfolg des Projekts. Damit hat das Projekt zumindest deutschlandweit eine Pionierrolle bei der Erstellung von virtuellen Forschungsumgebungen für die quantitativ-empirisch forschenden Sozial- und Wirtschaftswissenschaften und kann bei der Weiterentwicklung und Nutzung von VFUen in diesem Bereich durch die erbrachten Entwicklungen und Erfahrungen unterstützend wirken.

Die entwickelte Gesamtarchitektur der VFU erlaubt eine einfache konzeptuelle Anpassung für andere Fach-Communities oder Forschungsschwerpunkte und deren Workflows. Die entwickelte VFU soll daher künftig anderen Anwender/inne/n in einer erweiterten Testphase zur Verfügung gestellt werden. Dafür wird durch Eigenmittel der Projektpartner im Anschluss an die Projektförderung ein entsprechendes Betriebsmodell und eine Demoversion entwickelt.

Die im Projekt *VFU soeb 3* erreichten Ergebnisse, insbesondere die webbasierte Unterstützung kollaborativer Arbeit an Syntax sowie die kontrollierte Erfassung von Metadaten im Zusammenspiel mit den Suchfunktionen im Portal können die kollaborative Nutzung von Forschungsdaten sowie das vernetzte Arbeit in Verbundvorhaben der Sozialwissenschaften deutlich verbessern. Der zu erwartende Nutzen und die Akzeptanz dieser Lösungen hängen aber wesentlich davon ab, wie Verfahren des Fernrechnens mit Daten der FDZ in die VFU integriert werden können. Die Lösungen hierzu sind von großem Interesse für alle Forschungsvorhaben und –projekte, die Mikrodaten der Forschungsdateninfrastruktur nutzen. Durch das vorgelegte Proof of Concept zu Remote Data Access (siehe AP 4) hat der Verbund *VFU soeb 3* wichtige Grundlagen geschaffen, um zukünftige Entwicklungsarbeit in diesem Bereich anzustoßen. In einem möglichen Folgeprojekt will der Verbund schwerpunktmäßig die Realisierung von Remote Data Access innerhalb einer VFU verfolgen.

Darüber hinaus ist zu erwarten, dass die technischen Lösungen im Portalbereich sowie zur Anbindung der Middleware für andere Projekte bzw. VFU-Lösungen Vorbildcharakter haben



werden. Unter anderem hat die Universitätsmedizin Göttingen (UMG) Interesse an der Middleware und einigen zugehörigen Portlets zur Verwendung in medizinischen Projekten bekundet. Diese Komponenten wurden der UMG am 29.4.2013 vorgestellt. Neben der UMG könnte der Einsatz der Middleware auch für das INF-Projekt des SFB 963 (Forschungsdaten in der Astrophysik) relevant sein.

Das erstellte Konzept zur Langzeitarchivierung wurde auch vor dem Hintergrund einer potenziellen Nutzung durch weitere Anwenderkreise verfasst, welche die VFU für ihre Zwecke adaptieren wollen.

## **II.5 Fortschritt bei anderen Stellen**

Während der Durchführung des Vorhabens sind den Projektpartnern keine vergleichbaren Entwicklungen im Bereich der empirischen Sozialwissenschaften bekannt geworden. Das INF-Projekt des Bielefelder Sonderforschungsbereichs 882 (Friedhoff u.a.) verfolgt z.Tl. ähnliche Zielsetzungen, strebt jedoch dafür keine generischen, nachnutzbaren Lösungen an.

## **II.6 Veröffentlichungen und Transfer**

### *II.6.1 Veröffentlichungen*

Bartelheimer, Peter/ Cronjäger, Sarah/ Lehweß-Litzmann, René (2013): Eine virtuelle Forschungsumgebung für die Sozial- und Wirtschaftswissenschaften. SOFI-Mitteilungen. 7. Jahrgang. Ausgabe 19. 7-8.

Bartelheimer, Peter/ Cronjäger, Sarah (in Arbeit): Die Virtuelle Forschungsumgebung für die Sozioökonomische Berichterstattung (Arbeitstitel). Working-Paper beim RatSWD 2014. (SOFI, AP 1)

- Jensen, Uwe/ Schweers, Stefan (2014): Das erweiterte Metadatenschema der *VFU soeb 3* GESIS-Technical Reports – 2014/13. [http://www.etracker.de/lnkcnt.php?et=qPKGYYV&url=http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/gesis\\_methodenberichte/2014/TechnicalReport\\_2014-13.pdf&lnkname=fileadmin/upload/forschung/publikationen/gesis\\_reihen/gesis\\_methodenberichte/2014/TechnicalReport\\_2014-13.pdf](http://www.etracker.de/lnkcnt.php?et=qPKGYYV&url=http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2014/TechnicalReport_2014-13.pdf&lnkname=fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2014/TechnicalReport_2014-13.pdf); Stand: 01.10.2014. (GESIS, AP 7)
- Jensen, Uwe/ Schweers, Stefan/ Carevic, Zeljko (2014): Die Metadateneditoren der *VFU soeb 3*. GESIS-Technical Reports – 2014/14. [http://www.etracker.de/lnkcnt.php?et=qPKGYYV&url=http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/gesis\\_methodenberichte/2014/TechnicalReport\\_2014-14.pdf&lnkname=fileadmin/upload/forschung/publikationen/gesis\\_reihen/gesis\\_methodenberichte/2014/TechnicalReport\\_2014-14.pdf](http://www.etracker.de/lnkcnt.php?et=qPKGYYV&url=http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2014/TechnicalReport_2014-14.pdf&lnkname=fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methodenberichte/2014/TechnicalReport_2014-14.pdf). Stand: 01.10.2014. (GESIS, AP 7/8)
- Schiller, David/ Goebel, Jan/ Mika, Tatjana/ Pohl, Christof/ Boehme, Christian (in Arbeit): Entwicklung einer Arbeitsumgebung für Projektteams mit Zugang zu Mikrodaten für die Forschung (Arbeitstitel). Working-Paper beim RatSWD 2014. (FDZ IAB, FDZ SOEP, FDZ RV [AP 4], GWDG [AP 5/6])

## II.6.2 Vorträge

- Bartelheimer, Peter (2013): Daten in Virtuellen Forschungsumgebungen. Präsentation auf dem Symposium Forschungsdateninfrastrukturen des Helmholtz-Zentrums in Potsdam am 22.01.2013. [http://handytv.sofi-goettingen.de/fileadmin/Peter\\_Bartelheimer/Material/Vortraege/pb\\_FDI\\_2013\\_Forschungsdaten.pdf](http://handytv.sofi-goettingen.de/fileadmin/Peter_Bartelheimer/Material/Vortraege/pb_FDI_2013_Forschungsdaten.pdf), Stand: 02.10.2014 (SOFI, AP 1)
- Jensen, Uwe (2014): Einführung in das Metadatenschema der *VFU soeb 3*. Präsentation auf der Fachtagung: Das Portal, die Daten und wir - Eine Virtuelle Forschungsumgebung für die digitale Infrastruktur. Berlin 24.01.2014. [http://www.soeb.de/fileadmin/redaktion/downloads/VFU/VFU\\_Tagung\\_Jensen\\_final.pdf](http://www.soeb.de/fileadmin/redaktion/downloads/VFU/VFU_Tagung_Jensen_final.pdf). Stand: 04.10.2014. (GESIS, AP 7)
- Jensen, Uwe (2014): Metadata Requirements to document Data Analyses and Syntax Files in a Virtual Research Environment (VRE) – The use case *soeb 3*<sup>e</sup>. Presentation: EDDI 2014. 6th Annual European DDI User Conference. London. December 02.-03.2014. (Vortrag eingereicht und akzeptiert) (GESIS, AP 7)

- Mathiak, Brigitte (2014): A virtual research environment for syntax sharing. Presentation: International Association for Social Science, Information Services & Technology (IAS-SIST) 2014 in Toronto. 6W Research Environments. [http://www.library.yorku.ca/binaries/iassist2014/6W/2014\\_6W\\_Mathiak.ppt](http://www.library.yorku.ca/binaries/iassist2014/6W/2014_6W_Mathiak.ppt), Stand: 07.10.2014. (GESIS, AP 8)
- Mika, Tatjana/ Schiller, David (2013): Die Virtuelle Forschungsumgebung für die Sozioökonomische Berichterstattung. Präsentation auf der 10. Sitzung des ständigen Ausschusses FDI des RatSWD am 10.10.2013. (FDZ-RV, FDZ-IAB, AP 4)
- Pohl, Christof (2013): Liferay Portal as Platform for Virtual Research Environments. Presentation: IASSIST. 31.05.2013 Köln. (GWDG, AP 5)
- Pohl, Christof (2014): Die Architektur der Virtuellen Forschungsumgebung. Präsentation auf der Fachtagung: Das Portal, die Daten und wir - Eine Virtuelle Forschungsumgebung für die digitale Infrastruktur. Berlin 24.01.2014. [http://www.soeb.de/fileadmin/redaktion/downloads/VFU/VFU\\_Tagung\\_Pohl\\_final.pdf](http://www.soeb.de/fileadmin/redaktion/downloads/VFU/VFU_Tagung_Pohl_final.pdf). Stand: 04.10.2014. (GWDG, AP 5/6)
- Pohl, Christof/ Boehme, Christian (2014): Forschungsdatenmanagement mit CDSTAR am Beispiel der *soeb 3 VFU*. Präsentation: DV-Treffen der Max-Planck-Institute am 17. September 2014 in Berlin. (GWDG, AP 5/6)
- Schmidt, Tanja (2014): Ergebnisse aus zwei Nutzungsstudien. Präsentation auf der Fachtagung: Das Portal, die Daten und wir - Eine Virtuelle Forschungsumgebung für die digitale Infrastruktur. Berlin 24.01.2014. [http://www.soeb.de/fileadmin/redaktion/downloads/VFU/VFU\\_Tagung\\_Schmidt\\_final.pdf](http://www.soeb.de/fileadmin/redaktion/downloads/VFU/VFU_Tagung_Schmidt_final.pdf), Stand: 04.10.2014. (SOFI, AP 2)
- Schiller, David (2014): Weitergedacht – Remote Data Access in der Virtuellen Forschungsumgebung: Proof of Concept. Präsentation auf der Fachtagung: Das Portal, die Daten und wir - Eine Virtuelle Forschungsumgebung für die digitale Infrastruktur. Berlin 24.01.2014. [http://www.soeb.de/fileadmin/redaktion/downloads/VFU/VFU\\_Tagung\\_Schiller\\_final.pdf](http://www.soeb.de/fileadmin/redaktion/downloads/VFU/VFU_Tagung_Schiller_final.pdf), Stand: 02.10.2014. (FDZ-IAB, AP 4)

## II.7 Literatur

- Altenhöner, Reiner/ Oellers, Claudia (2012): Langzeitarchivierung von Forschungsdaten - Standards und disziplinspezifische Lösungen. ISBN 978-3-944417-00-4. SCIVERO Verlag. Berlin.
- Bartelheimer, Peter/ Schmidt, Tanja (2011): Modellprojekt „Kollaborative Datenauswertung und Virtuelle Arbeitsumgebung“ – VirtAug. Abschlussbericht. Download unter: [http://www.soeb.de/fileadmin/redaktion/downloads/VirtAug/VirtAug\\_Abschlussbericht.pdf](http://www.soeb.de/fileadmin/redaktion/downloads/VirtAug/VirtAug_Abschlussbericht.pdf), Stand: 10.10.2014.
- Data Documentation Initiative (DDI) Alliance (2014): DDI 2 - Codebook version; <http://www.ddialliance.org/Specification/DDI-Codebook/>, Stand: 06.10.2014.
- Deutsche Forschungsgemeinschaft (DFG) (1998): Gute wissenschaftliche Praxis. Denkschrift. Weinheim. Download: [http://www.dfg.de/download/pdf/dfg\\_im\\_profil/reden\\_stellungnahmen/download/empfehlung\\_wiss\\_praxis\\_0198.pdf](http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_0198.pdf), Stand: 02.10.2014.
- Deutsche Forschungsgemeinschaft (DFG) (2010): Informationsverarbeitung an Hochschulen – Organisation, Dienste und Systeme. Empfehlungen der Kommission für IT-Infrastruktur 2011-2015. Bonn. Download: [http://www.hof.uni-halle.de/dateien/pdf/DFG\\_Empfehlungen\\_IT\\_2011\\_2015.pdf](http://www.hof.uni-halle.de/dateien/pdf/DFG_Empfehlungen_IT_2011_2015.pdf), Stand: 08.10.2014.
- Deutsche Forschungsgemeinschaft (DFG) (2013): Vernetzung – Austausch – Nachnutzung. Workshop zur Präsentation von Projekten aus dem Förderbereich Informationsmanagement der DFG. Bonn.
- Dickmann, Frank/ Enke, Harry/ Harms, Patrick (2010): Technische Evaluation der Grid-Technologie für das Modellprojekt Kollaborative Datenauswertung und virtuelle Arbeitsumgebung – VirtAug. soeb-Arbeitspapier 2010-1. Download: [http://www.soeb.de/fileadmin/redaktion/downloads/VirtAug/Expertise\\_VirtAug.pdf](http://www.soeb.de/fileadmin/redaktion/downloads/VirtAug/Expertise_VirtAug.pdf); Stand: 07.10.2014.
- Friedhoff, Stefan/ Meier zu Verl, Christian/ Pietsch, Christian/ Meyer, Christian/ Vompras, Johanna/ Liebig, Stefan (2013): Social Research Data. Documentation, Management, and Technical Implementation within the SFB 882. SFB 882 Working Paper Series, 16. Bielefeld: DFG Research Center (SFB) 882 From Heterogeneities to Inequalities. urn:nbn:de:0070-pub-25600355
- Gentzsch, Wolfgang (2007): Grid-Computing und die deutsche D-Grid-Initiative. In: Neuroth, Heike/Gentzsch, Wolfgang: 9-11.
- German Data Forum (RatSWD) (Hrsg.) (2010): Building on Progress. Expanding the Research Infrastructure for the Social, Economic, and Behavioral Sciences (Vol. I, II).
- Kommission Zukunft der Informationsinfrastruktur (zitiert als KZII) (2011): Gesamtkonzept für die Informationsinfrastruktur in Deutschland. Empfehlungen der Kommission Zukunft der Informationsinfrastruktur im Auftrag der Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder. April 2011. Download: [http://www.leibniz-gemeinschaft.de/fileadmin/user\\_upload/downloads/Infrastruktur/KII\\_Gesamtkonzept.pdf](http://www.leibniz-gemeinschaft.de/fileadmin/user_upload/downloads/Infrastruktur/KII_Gesamtkonzept.pdf), Stand: 07.10.2014.
- Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (KVI) (Hrsg.) (2001): Wege zu einer besseren informationellen Infrastruktur; Gutachten der vom Bundesministerium für Bildung und Forschung eingesetzten Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik. Baden-Baden: Nomos.
- Ludwig, Jens/ Enke, Harry (2013): Leitfaden zum Forschungsdaten-Management. Handreichungen aus dem WissGrid-Projekt. Glückstadt: Hülsbusch. Download: [http://webdoc.sub.gwdg.de/univerlag/2013/leitfaden\\_DGRID.pdf](http://webdoc.sub.gwdg.de/univerlag/2013/leitfaden_DGRID.pdf), Stand: 06.10.2014.
- Mochmann, Ekkehard (2010): e-Infrastructure for the Social Sciences. In: German Data Forum (RatSWD): 265-285.
- Neuroth, Heike/Gentzsch, Wolfgang (Hrsg.) (2007): Die D-Grid-Initiative. Göttingen: Universitätsverlag Göttingen.
- Neuroth, Heike/Lohmeier, Felix/Smith, Kathleen M. (2010): e-Humanities – eine virtuelle Forschungs-umgebung für die Geistes-, Kultur- und Sozialwissenschaften. In: BIBLIOTHEK Forschung und Praxis. Band 31, Heft 3, Seiten 272–279, ISSN (Print) 0341-4183, DOI: 10.1515/BFUP.2007.272, April 2008
- Pohl, Christof (2014): Die Architektur der Virtuellen Forschungs-umgebung. Präsentation auf der Fachtagung: Das Portal, die Daten und wir - Eine Virtuelle Forschungs-umgebung für die digitale Infra-

struktur. Berlin 24.01.2014. [http://www.soeb.de/fileadmin/redaktion/downloads/VFU/VFU\\_Tagung\\_Pohl\\_final.pdf](http://www.soeb.de/fileadmin/redaktion/downloads/VFU/VFU_Tagung_Pohl_final.pdf). Stand: 04.10.2014.

Projekt RADIESCHEN Rahmenbedingungen einer disziplinübergreifenden Forschungsdateninfrastruktur (2013): Report „Synthese“. Entspricht dem Report D6.3 „Abschlussbericht des Projekts und Roadmap für eine Infrastruktur für Forschungsdaten in Deutschland“ nach Projektantrag. 30.04.2013. DOI: [http://doi.org/10.2312/RADIESCHEN\\_007](http://doi.org/10.2312/RADIESCHEN_007), <http://gfzpublic.gfz-potsdam.de/pubman/item/escidoc:117053:3>.

Schmitt, Oliver/ Siemon, Andreas/ Schwardmann, Ulrich/ Hellkamp Mrcel (2014): GWDG Object Storage and Search Solution for Research: Common Data Storage Architecture (CDSTAR) - GWDG Bericht. Göttingen. Download: <http://www.gwdg.de/fileadmin/inhaltsbilder/Pdf/Publikationen/GWDG-Berichte/gwdg-bericht-78.pdf>, Stand: 08.10.2014.

Winkler-Nees, Stefan (2011): Anforderungen an wissenschaftliche Informationsinfrastrukturen. Working Paper Series des Rates für Sozial- und Wirtschaftsdaten 180. Download: [http://www.ratswd.de/download/RatSWD\\_WP\\_2011/RatSWD\\_WP\\_180.pdf](http://www.ratswd.de/download/RatSWD_WP_2011/RatSWD_WP_180.pdf), Stand: 07.10.2014.