

# Antrag im SAW-Verfahren 2014 der Leibniz-Gemeinschaft

Förderlinie 2 Nationale und internationale Vernetzung

## Titel

SowiDataNet – Forschungsdatenverbund für Sozial- und Wirtschaftswissenschaften: Infrastruktur für die eigenständige Archivierung und Distribution quantitativer Forschungsdaten der Sozial- und Wirtschaftswissenschaften

## Projektziele

*“There is growing awareness of the need for data storage and data preservation and a growing sense of urgency concerning this issue. All stakeholders (funding agencies, data producers, data consumers, data centres) agree, for various reasons, that something needs to be done to improve the current situation. At the same time, no one solution has been found.”* (Feijen, 2011: 29)

Ziel des Projekts SowiDataNet ist der Aufbau eines Forschungsdatenverbunds für Sozial- und Wirtschaftswissenschaften. Kern des Verbunds wird eine web-basierte, eigenständige Infrastruktur sein, welche eine niederschwellige Selbst-Archivierung, Dokumentation und Distribution von Forschungsdaten ermöglicht und sich dabei am konkreten Bedarf der Scientific Community orientiert. Im Mittelpunkt stehen dabei quantitative Daten aus den Sozial- und Wirtschaftswissenschaften und damit aus zwei besonders empirisch orientierten großen Wissenschaftsdisziplinen.

Der Begriff „Forschungsdaten“ wird weit gefasst. Er umfasst sowohl Primärdaten, die aufgrund der Nutzung empirischer Forschungsinstrumente neu produziert werden, als auch Sekundärdaten, die z.B. aus vorhandenen Quellen in spezifische neue Datensätze oder Datenbanken transformiert werden. Darüber hinaus umfasst der Begriff ebenfalls die Kombination (und evtl. Transformation) von Primär- und Sekundärdaten in neue Datensätze sowie die für Datentransformationen und -analysen benutzten Routinen oder Skripte.

Zielgruppe von SowiDataNet sind zunächst Projekte mit fehlenden oder geringen personellen und finanziellen Ressourcen für die Dokumentation und Archivierung von Forschungsdaten. Dies sind in der Regel mittlere und kleine Forschungs- oder Promotionsprojekte. Im Verlauf des Projekts ist ebenfalls geplant zu prüfen, ob und inwieweit Forschungsdaten aus größeren Projekten und Institutionen sowie etablierte Datenbestände in den aufzubauenden Verbund sinnvoll einzubinden sind. Eine Integration von Datenbeständen aus unterschiedlichen Quellen hat den Vorteil, dass Forschungsdaten der Sozial- und Wirtschaftswissenschaften in Deutschland nach einheitlichen Standards archiviert und dokumentiert werden, wobei zugleich individuelle Besonderheiten berücksichtigt werden können (z.B. Zugriffsschutz, Rechteverwaltung). Die Integration unterschiedlicher Quellen von Forschungsdaten führt zudem auch zu einer zentralen Recherchemöglichkeit für Forschungsdaten und wird, wo umsetzbar, einen direkten Zugang zu diesen Forschungsdaten ermöglichen.

Im Zuge des Projekts werden modellhaft am Datenlebenszyklus orientierte Workflows erarbeitet, die eine systematische Nutzung der im Projektkontext entwickelten technischen Infrastruktur für die Sicht- und ggf. Verfügbarmachung institutionell erzeugter Forschungsdaten erleichtern. Im Mittelpunkt soll dabei eine einfache und niedrigschwellige Möglichkeit der Datenproduzenten<sup>1</sup> stehen, Forschungsdaten zu speichern und dauerhaft zu archivieren. Dabei muss gewährleistet werden, dass die Datenproduzenten die Entscheidung über eine Sichtbarmachung ihrer Datensätze, das Rechtemanagement und die verwendeten Nutzungslizenzen selbst treffen können (vgl. Feijen 2011; McCullough 2009).

---

<sup>1</sup> Im Folgenden wird aus Gründen der besseren Lesbarkeit meist die männliche Schreibweise verwendet. Selbstverständlich ist die weibliche Form jeweils auch impliziert.

Gesichert ist weiterhin eine Beteiligung von außeruniversitären und universitären Institutionen an der Anforderungsanalyse, am Aufbau von SowiDataNet und an der Vermittlung in die Wissenschaftsgemeinschaft (siehe dazu unter „Projektstruktur“ die Institutionen, von denen ein Letter of Intent vorliegt). Dies trägt bereits in der Projektphase zu einer breiten Vernetzung von Personen, Projekten und Institutionen bei.

Durch die Entwicklung zentraler Community-Dienste im Rahmen von SowiDataNet werden außerdem die wissenschaftlichen Anbieter von Daten mit den Nutzern der Daten über eine gemeinsame Infrastruktur stärker vernetzt. Das Angebot von Community-Diensten soll eine intensive Interaktion der beteiligten Datenanbieter und Datennutzer von SowiDataNet anregen. Solche Dienste ermöglichen etwa eine Bewertungsfunktion für Angebote anderer Nutzer, Diskussionsforen oder eine Einbindung in Social-Media-Netzwerke, die zur aktiven Teilnahme aller Wissenschaftler anregen. Neben den im Zusammenhang mit dem Begriff „Web 2.0“ oft genannten „Social Networking“-Werkzeugen sind dabei auch die „Social Awareness“-Tools zu berücksichtigen (siehe McMahon et al. 2012).

Mit dem Projekt werden die Möglichkeiten zur Aufbewahrung aller nationalen Forschungsdaten aus den Sozial- und Wirtschaftswissenschaften substanziell verbessert. Perspektivisch sollen dadurch eine effektive Langzeitarchivierung (beispielsweise durch Anbindung an das Datenarchiv der GESIS), eine standardisierte Dokumentation und eine effektive Zugänglichkeit von Forschungsdaten deutlich vereinfacht werden. Aus diesem Punkt leitet sich sowohl die hohe strategische Bedeutung des Projekts für GESIS als auch für die Sozial- und Wirtschaftswissenschaften insgesamt ab.

Indem dezentral erzeugte Forschungsdaten unter Nutzung einer gemeinsamen Infrastruktur miteinander vernetzt, archiviert, beschrieben, zitierbar und möglichst auch zugänglich gemacht werden, ergeben sich auch verbesserte Möglichkeiten der Daten-Reanalyse. Ferner ist für die Interpretierbarkeit der Forschungsdaten eine Kontaktmöglichkeit zu den Primärforschern, wie sie durch SowiDataNet in Form eines sozialen Netzwerks entsteht, eine wertvolle Hilfe. Gleichzeitig hilft die Anbindung an bestehende Langzeitarchivierungsdienste dabei, die Daten so zu dokumentieren, dass sie später auch ohne Kontakt zu den Primärforschern verständlich sind. Insgesamt zielt das Projekt somit ab auf den Aufbau und die Etablierung einer persistenten Forschungsdateninfrastruktur für die Sozial- und Wirtschaftswissenschaften und wirkt einer möglichen Zerklüftung der Forschungsinfrastruktur aktiv entgegen.

## **Hintergrund**

Forschungsdaten sind in vielen Forschungsdisziplinen die Grundlage für wissenschaftliche Erkenntnisse und daraus resultierende Veröffentlichungen. Durch Replikation der Analysen wird ein Beitrag zur Erhöhung der Transparenz von Forschungsprozessen geleistet: Replikation ermöglicht Sicherung und weitere Entwicklung des Wissens in einem Fachgebiet und ist darüber hinaus eine nützliche Basis für Ausbildung in der Wissenschaft (vgl. King 1995). Für die Durchführung von Replikationsstudien müssen Forschungsdaten jedoch zugänglich sein. Ihre freie und langfristige Verfügbarkeit sichert den Erkenntnisgewinn und stellt die notwendige Transparenz und Nachvollziehbarkeit im Forschungsprozess sicher (vgl. King 1995; Jasny et al. 2011). Das aktive Nachnutzen von Forschungsdaten sorgt darüber hinaus für die Vermeidung von redundanten Untersuchungen, für eine fachliche Qualitätskontrolle und kann zu stetig aktualisierten Forschungsdatenbeständen führen. Ferner trägt die Verfügbarkeit von Forschungsdaten auch dazu bei, dass die Primärforscher öfter zitiert werden (vgl. Piwowar et al. 2007), womit ein wesentlicher Anreiz für die oft zeitaufwändige Dokumentation der Daten gegeben wird.

Viele Institutionen haben eine Empfehlung für die Archivierung von Forschungsdaten ausgesprochen. So heißt es z.B. in den Richtlinien der DFG: *„Primärdaten als Grundlagen für Veröffentlichungen sollen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, für zehn Jahre aufbewahrt werden.“* (DFG 1998: 12) Der Rat für Sozial- und Wirtschaftsdaten hat in seinem Bericht „Auf Erfolge aufbauend“ die Notwendigkeit eines Angebots zur Selbstarchivierung betont und die Verwendung von Metadaten-Standards

empfohlen (vgl. RatSWD 2011: 32). Auch der Wissenschaftsrat hat in seinen Empfehlungen neben der Datendokumentation und -aufbereitung vor allem die Forderung nach besserer Zugänglichkeit von Forschungsdaten erhoben (Wissenschaftsrat 2012: 54).

Für eine Umsetzung dieser Empfehlungen der Wissenschaftsorganisationen existiert jedoch bislang für die Sozial- und Wirtschaftswissenschaften keine adäquate Infrastruktur. An dieser Stelle setzt das Projekt SowiDataNet an. Die geplante Infrastruktur folgt damit auch der Vision, die die EU im Report „Riding the wave“ beschrieben hat: *„All of these principles – our vision - point in the direction of an infrastructure that supports seamless access, use, reuse, and trust of data. It suggests a future in which the data infrastructure becomes invisible, and the data themselves have become infrastructure – a valuable asset, on which science, technology, the economy and society can advance.“* (EU 2010: 24)

### **Zielgruppen und Reichweite**

Zielgruppe ist die sozial- und wirtschaftswissenschaftliche Forschung in Deutschland, wobei zwischen Datenanbietern und Datennutzern zu differenzieren ist.

*Datenanbieter* sollen über SowiDataNet ihre Forschungsdaten in einfacher Weise dokumentieren, archivieren und anderen zugänglich machen können. Dabei richtet sich SowiDataNet zunächst an kleine bis mittlere Projekte mit Forschungsdaten, z.B. Promotionsprojekte. Die Anbindung von Graduiertenkollegs, Sonderforschungsbereichen oder Experimentallaboren wird ebenfalls im Zuge des Projektes eruiert, wobei es sich zumindest bei den ersten beiden Kategorien um Vorhaben handelt, die zum Teil über eigene Infrastrukturen verfügen. Zu einem späteren Zeitpunkt können darüber hinaus auch Forschungsdaten größerer Projekte und Institutionen in den Verbund integriert werden, um zentrale Recherchemöglichkeiten zu schaffen.

*Datennutzer* sollen über SowiDataNet in den Metadaten recherchieren, Daten für Reanalysen oder neue Forschungsvorhaben nutzen können und einen möglichst zentralen Zugang für die Arbeit mit den Forschungsdaten erhalten. Ihre Erfahrungen dabei werden über die Community-Dienste anderen Nutzern und Anbietern zur Verfügung stehen.

Mit den adressierten Disziplinen – Sozial- und Wirtschaftswissenschaften – werden Fächer angesprochen, die zusammen mit großem Abstand die meisten Studierendenzahlen und dazu entsprechende Lehrstuhl-Strukturen in Deutschland auf sich vereinen (siehe DESTATIS 2012). Vor diesem Hintergrund wird das Projekt mit einer innovativen Forschungsinfrastruktur zwei überdurchschnittlich große Zielgruppen unterstützen. Angestrebt wird nach Ende der Projektlaufzeit eine durch die modulare Struktur mögliche Erweiterung auf angrenzende Fachdisziplinen (z.B. Psychologie), welche die dortigen fachspezifischen Besonderheiten (z.B. Metadaten) berücksichtigt.

Das Projekt ist zugeschnitten auf den Bereich der *quantitativen* Forschungsdaten, da die Projektpartner hier über eine weitreichende Expertise verfügen. Das Handling *qualitativer* Forschungsprimärdaten (z.B. als Audio-/Videodateien) stellt ein Problem sui generis dar (siehe Knoblauch et al. 2011) und wird in anderweitigen Kooperationen der Partner (GESIS, WZB) etwa mit dem Bremer Archiv für Lebenslaufforschung und dem Deutschen Institut für Internationale Pädagogische Forschung weiterverfolgt.

### **Arbeitspakete**

Der operative Aufbau der Infrastruktur wird auf die konkreten Bedürfnisse der Wissenschaft abgestimmt. Gegenstand der **Anforderungsanalyse (AP1)** wird daher die Ermittlung von Kontexten sein, in denen relevante Forschungsdaten entstehen (z.B. Promotionsvorhaben, ggf. Auftragsforschung, Publikationen in Zeitschriften, etc.). Ebenso werden die von den Forschern verwendeten Datei- und Metadatenformate ermittelt. Barrieren, die einer Publikation von Forschungsdaten durch die Wissenschaftler entgegenstehen (z.B. Urheberrechte, datenschutzrechtliche Probleme oder Restriktionen bei gekauften oder lizenzierten Datensätzen) werden ermittelt, um sie für die zukünftige Infrastruktur möglichst zu reduzieren.

Weiterhin wird untersucht, welche Funktionsmerkmale die Infrastruktur aufweisen muss, um die Sichtbarmachung von Daten in anderen Kontexten zu ermöglichen (z.B. Nachweis der Daten in zentralen Katalogen, Verfügbarkeit der Daten als Linked Open Data, Vergabe von persistenten Identifikatoren für die Daten, Verknüpfung von Daten mit Publikationen). Zudem wird geprüft, welche Community-Dienste zur Vernetzung der Datenanbieter mit den Datennutzern auf einer gemeinsamen Plattform sinnvoll angeboten werden können (z.B. Diskussionsforen, Bewertungs- oder Benachrichtigungsfunktionen). Die Anforderungsanalyse vollzieht sich in einem mehrstufigen Verfahren: Sie beginnt mit der Literatursichtung und der Einbeziehung der Ergebnisse anderer, ähnlich gelagerter Forschungsprojekte (z.B. Steinhart et al. 2011 oder Feijen 2011). Insbesondere werden hier auch die Ergebnisse der Untersuchungen zu Anreizsystemen im Rahmen des EDaWaX-Projekts der ZBW herangezogen (Siegert et al. 2012). Im Anschluss ist eine eigene Erhebung geplant. Bei der Anforderungsanalyse erfolgt eine Unterstützung durch Institutionen aus dem Bereich der Hochschulen und der außeruniversitären Forschungsinstitutionen, um so eine Rückkopplung an die Bedarfe der Forschungspraxis sicherzustellen. Dies geschieht vor allem durch Workshops und qualitative Interviews gewährleistet. Der möglichst frühzeitig in AP4 entwickelte Prototyp wird mit diesen Nutzern fortwährend getestet und mit deren Anforderungen abgeglichen.

Im Arbeitspaket **Organisationsmodell (AP2)** geht es um das Zusammenwirken der beteiligten Akteure sowie ihre jeweilige Rolle und ihren Beitrag zu SowiDataNet. Die Richtlinien zur Nutzung der Infrastruktur werden auf Basis der Anforderungsanalyse (AP1) und in Abstimmung mit allen Beteiligten festgelegt. Dies beinhaltet u.a. auch rechtliche Aspekte bzgl. der Gewährung von Nutzungsrechten für SowiDataNet sowie ggf. von den Nutzern einzuräumende spezifische Rechte, die im Kontext der Langzeitarchivierung der Daten notwendig sind. Das erarbeitete Organisationsmodell wird auf seine langfristige, über die Projektlaufzeit hinausreichende Tragfähigkeit untersucht mit dem Ziel, ein dauerhaft tragfähiges Geschäftsmodell zu erarbeiten, welches das Organisationsmodell auch finanziell nachhaltig sichert.

Teil des AP2 ist darüber hinaus die Erarbeitung von Workflows zur Unterstützung der Forschenden bei der Speicherung und der Dokumentation von Daten in SowiDataNet in den verschiedenen Phasen des Datenlebenszyklus. Zu diesem Zweck wird ein strukturiertes Vorgehen entwickelt, mit dessen Hilfe Fragen des Forschungsdatenmanagements durch spezialisierte Mitarbeiter in den Instituten operativ bearbeitet werden können (z.B. durch Gesprächsleitfäden). Exemplarische, möglichst standardisierte Workflows für die Nutzung der Infrastruktur in einem institutionellen Kontext werden anhand konkreter Fallbeispiele entwickelt. Diese orientieren sich an den notwendigen Funktionen und Inhalten, wie sie in der Anforderungsanalyse erhoben worden sind.

Die Erarbeitung von adäquaten **Metadaten-Strukturen (AP3)** und die Dokumentation nach anerkannten Standards sind neben den technischen Projektkomponenten gleichermaßen wesentliche Projektbestandteile. Dabei werden sowohl aktuelle sozial- und wirtschaftswissenschaftliche Datendokumentations-Standards wie DDI und SDMX berücksichtigt als auch relevante Standards für Bibliothekswesen und Archivierung, etwa Dublin Core und PREMIS (vgl. Jensen et al. 2011; Zenk-Möltgen 2012). Ebenso werden bereits in Projektkontexten eingesetzte Schemata wie z.B. das da|ra Metadatenschema (siehe Hausstein et al. 2011a) oder das DBK Metadatenschema (siehe Zenk-Möltgen 2012) einbezogen. In diesem Zusammenhang finden auch Schnittstellen zu Persistent-Identifizierungssystemen, dem Social und dem Semantic Web Berücksichtigung. Die so definierten Metadaten müssen im Übrigen anschlussfähig an existierende Repositorien sein und auch den Bereich der geplanten Community-Dienste (Bewertungsfunktionen, Foren, Social Media usw.) abdecken. Eine gemeinsame Grundlage wird entwickelt und dient als Basis für die Softwareentwicklung.

Die **Software-Entwicklung (AP4)** für SowiDataNet zielt auf ein zentrales Webportal, in dem die Datenanbieter die Archivierung und die Dokumentation nach anerkannten Standards durchführen können. Darüber hinaus sollen eine lokale Sicht auf die eigenen

Forschungsdaten und ein differenziertes Rechtemanagement ermöglicht werden. Dabei werden die Ergebnisse der Arbeitspakete zu Anforderungsanalyse, Organisationsmodell und Metadaten berücksichtigt. Sollte sich im Rahmen der Anforderungsanalyse zeigen, dass Interesse an einer dezentralen Lösung für eine lokale Vorbereitung der Archivierung besteht, wird dies im Rahmen der Architekturkonzeption entsprechend zu berücksichtigen sein.

Über das SowiDataNet-Portal werden eine zentrale Suche und der Datenzugang für die Scientific Community angeboten. Zusätzlich erfolgt die Anbindung an Dienste für eine Registrierung der Daten mit dauerhaften Identifikatoren und für eine langfristige Sicherung der Daten. Die dann mögliche, flächendeckende Zitation von Forschungsdaten etwa nach den von DataCite vorgeschlagenen Standards (siehe DataCite 2011) führt zu einer erhöhten Sichtbarkeit dieser Forschungsdaten.

Die aufzubauende Infrastruktur soll einerseits eine möglichst niedrighschwellige Einspeisung *lokal generierter* Forschungsdaten in eine *zentrale* Forschungsdateninfrastruktur ermöglichen, wobei Fragen der Versionierung, Langzeitarchivierung, Nutzungsstatistiken und einer Staffelung von Zugriffsrechten angemessen zu berücksichtigen sind. Andererseits soll auch die dezentrale Sichtbarmachung von Forschungsdaten über eine auf den Datenanbieter begrenzte Sicht auf die Datenbestände ermöglicht werden, indem z.B. entsprechende Such- und Anzeigefunktionen nahtlos in den Webauftritt eines institutionellen Anbieters integrierbar sind („lokale Vitrine“). Somit werden Anreize für einzelne Einrichtungen geschaffen, die Infrastruktur für ihre Daten zu nutzen. Für Institutionen, an denen bereits lokale Forschungsdatenrepositorien bestehen oder im Aufbau sind, sollen auch Schnittstellen (etwa basierend auf DDI oder SDMX) für die möglichst simultane Einspeisung der Forschungsdaten in die Infrastruktur mit dem geringstmöglichen Aufwand für den Forscher entwickelt werden, wie z.B. im Fall von DataStaR, (siehe Steinhart 2011).

Eine modulare Architektur der SowiDataNet-Infrastruktur soll die Erweiterbarkeit der Software nach der Projektlaufzeit ermöglichen. Die Software soll aus wiederverwendbaren Software-Bausteinen bestehen, mit einer Open-Source-Lizenz veröffentlicht werden und daher idealerweise auf einem bestehenden Open-Source-Framework aufsetzen. Möglichst frühzeitig wird ein Prototyp entwickelt, der zum einen das Grundmodell veranschaulicht und zum anderen die Anforderungspunkte in AP1 besser zu identifizieren hilft.

Die technische Infrastruktur für SowiDataNet soll durch eine Evaluation bestehender Software ausgewählt werden, wobei sowohl eigene Entwicklungen wie etwa da|ra (siehe Hausstein et al. 2011b) als auch Repository-Software einbezogen werden sollen. Die Projektpartner verfügen bereits über Erfahrungen mit Repository-Software, welche bei den Angeboten von EconStor, SSOAR und Datorium eingesetzt wird (vgl. „Stand des Wissens“). Zu den Lösungen, die international verbreitet sind, zählen DSpace, EPrints, Fedora, MyCoRe, und OPUS, CKAN, Git, Mendeley oder Dataverse (siehe Open Society Institute 2004; Kunze 2012; Loeliger 2006; Zaugg et al. 2011; Crosas 2011). Diese Systeme bieten jeweils verschiedene Funktionalitäten und Lizenzmodelle sowie unterschiedlich stark ausgeprägte Entwickler- und Nutzer-Communities. Die vorhandene Expertise der Partner GESIS und ZBW ermöglicht es, anhand der im Projekt erhobenen Anforderungen zügig eine Entscheidung für ein System zu treffen.

Auf Basis der Ergebnisse der Anforderungsanalyse (AP1) sowie der in AP2 entwickelten Workflows werden **Beispielinhalte (AP5)** für die Infrastruktur zur systematischen Generierung der entsprechenden Metadaten durch die beteiligten Partner erarbeitet. Dabei können die Projektpartner auf die in ihren Instituten vorhandenen Forschungsdaten zurückgreifen. Damit sollen die Leistungsfähigkeit der Infrastruktur für verschiedene Datentypen gezeigt und eine Grundlage für den Weiterbetrieb nach der Projektlaufzeit geschaffen werden. So sollen weitere Forscher sowie Institutionen für eine Beteiligung am SowiDataNet-Portal gewonnen werden.

Der **Transfer (AP6)** der entwickelten Konzepte und Infrastrukturkomponenten im Sinne einer Vermittlung an die Scientific Community wird durch ein Schulungskonzept ermöglicht. Schulungsmaterialien wie eine ausführliche Dokumentation, Präsentationen, Broschüren und

How-to-Anleitungen werden erstellt. Insbesondere ist geplant, die Schulungsmaterialien über ein zentrales Blog bereitzustellen und die Materialien über einschlägige Publikationen und die Nutzung von Social-Media-Diensten in der Community bekannt zu machen. Dazu gehören auch das Marketing bei den Anwender-Zielgruppen und der Besuch von einschlägigen Konferenzen.

Das **Projektmanagement (AP7)** wird von GESIS durchgeführt. Dazu gehören die Projektablaufplanung und -steuerung und die Kommunikation unter den beteiligten Partnern sowie den unterstützenden Institutionen. Die Organisation der Zusammenarbeit in den Arbeitspaketen wird über regelmäßige Telefon- und Videokonferenzen sowie Arbeitstreffen und die geplanten Workshops erfolgen. Eine wichtige Rolle für die Entwicklung des Forschungsdatenverbunds SowiDataNet spielt die Unterstützung durch Institutionen aus dem Bereich der Hochschulen und der außeruniversitären Forschungsinstitutionen. Sie werden zu Workshops eingeladen, die im Rahmen der Anforderungsanalyse und zur Zwischenevaluierung von Projektergebnissen durchgeführt werden.

Zusammenfassend ergeben sich damit für die Konzeption und Entwicklung des Forschungsdatenverbundes SowiDataNet die folgenden Arbeitspakete: AP1 Anforderungsanalyse, AP2 Organisationsmodell, AP3 Metadaten-Strukturen, AP4 Software-Entwicklung, AP5 Beispielinhalte, AP6 Transfer und AP7 Projektmanagement.

### **Projektstruktur**

Das Projekt insgesamt wird mit Blick auf die Gewährleistung einer effektiven und effizienten Governance getragen von vier Partnern, nämlich GESIS, DIW Berlin, WZB und ZBW. Dabei ergänzen sich die vier Partner in sinnvoller und komplementärer Weise: Bei den beteiligten Projektpartnern handelt es sich um je zwei Einrichtungen mit überwiegendem *Forschungsschwerpunkt* und solche mit überwiegendem *Infrastrukturschwerpunkt*. Dies ist eine wesentliche Voraussetzung für die Entwicklung einer an den Bedarfen und methodologischen Standards der Sozial- und Wirtschaftswissenschaften orientierten Infrastruktur. Nur so können die Erfassung mittlerer und kleiner Projekte auf eine hinreichend breite Basis gestellt und eine an den realen Bedarfen der Forschung orientierte Infrastruktur-Konzeption ermöglicht werden.

GESIS und ZBW nehmen im Projekt die Rolle als Infrastrukturentwickler und -anbieter für Sozial- und Wirtschaftswissenschaften ein. GESIS und DIW Berlin spielen überdies eine Doppelrolle: Beide führen selbst große Studien durch (Allgemeine Bevölkerungsumfrage ALLBUS, Sozio-oekonomisches Panel SOEP). GESIS und DIW Berlin erheben aber auch ständig weitere kleinere Surveys und generieren Forschungsdaten, z.B. im Rahmen von Fallstudien. Im Projekt soll deutlich getrennt werden zwischen ihrer Rolle als Datenanbieter und der als Anbieter von Dienstleistungen für andere Datenanbieter im Rahmen des SowiDataNet-Verbunds. GESIS und ZBW werden die technischen Entwicklungsarbeiten des Projekts übernehmen.

GESIS wird als Hauptantragsteller für die Fortführung der Infrastruktur nach Ende der Projektlaufzeit aus eigenen Mitteln verantwortlich sein. Die Zusammenarbeit der Partner wird nach der Projektlaufzeit durch eine laufende Nutzung von SowiDataNet durch die Forschenden an den Einrichtungen fortgeführt. Ebenso wird die Vernetzung der Wissenschaftler aus den universitären und außeruniversitären Forschungseinrichtungen durch SowiDataNet weiter vertieft.

Das DIW Berlin und das WZB fungieren als Datengeber von insgesamt sehr heterogenen sozial- und wirtschaftswissenschaftlichen Forschungsdaten, indem sie in zahlreichen Projekten mit weiteren Partnern zusammen Forschungsdaten erzeugen. Sie bringen darüber hinaus in das Projekt eigene Datenmanagement-Erfahrungen ein.

Eine wichtige Rolle für die Entwicklung des Forschungsdatenverbunds SowiDataNet spielt die Unterstützung durch Institutionen aus dem Bereich der Hochschulen und der außeruniversitären Forschungsinstitutionen. Ein Letter of Intent zur Unterstützung des Projekts liegt bereits von folgenden Institutionen vor: Bamberg Graduate School of Social

Sciences (BAGSS); Bremen International Graduate School of Social Sciences (BIGSSS); Cologne Graduate School in Management, Economics and Social Sciences (CGS); Datenservicezentrum Betriebs- und Organisationsdaten an der Universität Bielefeld (DSZ-BO); Deutsches Institut für Internationale Pädagogische Forschung (DIPF); GIGA German Institute of Global and Area Studies Leibniz-Institut für Globale und Regionale Studien; Humboldt-Universität zu Berlin (Computer- und Medienservice); Sonderforschungsbereich „Von Heterogenitäten zu Ungleichheiten“ (SFB 882); Universitäts- und Stadtbibliothek Köln; Universität Mannheim (inkl. Graduate School of Economic & Social Sciences GESS). Sie werden durch ihre Beteiligung an der Anforderungsanalyse und die Zwischenevaluierung von Projektergebnissen auf Workshops zu einer praxisgerechten Entwicklung und Implementierung des Forschungsdatenverbunds SowiDataNet beitragen. Zudem ist eine enge Zusammenarbeit mit dem Rat für Sozial- und Wirtschaftsdaten (RatSWD) während der gesamten Laufzeit des Projekts vorgesehen: Einer der Antragssteller, Gert G. Wagner, ist Mitglied des RatSWD und zur Zeit dessen Vorsitzender. Damit ist gewährleistet, dass die im Projekt zu entwickelnde Infrastruktur anschlussfähig zu anderen Forschungsdateninfrastrukturen in den Sozial- und Wirtschaftswissenschaften ist (Forschungsdatenzentren und Datenservicezentren).

Die Aufgabenverteilung im Projekt ist wie folgt vorgesehen:

Arbeitspaket	GESIS	DIW Berlin	WZB	ZBW
AP1 Anforderungsanalyse	Mitarbeit	Mitarbeit	Leitung	Mitarbeit
AP2 Organisationsmodell		Mitarbeit	Mitarbeit	Leitung
AP3 Metadaten-Strukturen	Leitung		Mitarbeit	
AP4 Software-Entwicklung	Leitung			Mitarbeit
AP5 Beispielinhalte		Mitarbeit	Leitung	
AP6 Transfer	Mitarbeit	Leitung	Mitarbeit	Mitarbeit
AP7 Projektmanagement	Leitung			

### Ablaufplan

Geplant ist, das Projekt mit der Anforderungsanalyse (AP1), der Entwicklung eines nachhaltigen Geschäftsmodells für die Dienstleistung (AP2) und der Ausarbeitung der Metadaten-Strukturen (AP3) zu beginnen (s. Abb. unten). AP1 wird nach 18 Monaten abgeschlossen sein, AP2 nach 21 Monaten und AP3 nach 30 Monaten. Ab dem 7. Monat kann mit ersten Softwareentwicklungen begonnen werden (AP4), die bis Projektende fortgeführt werden. Ab dem 13. Monat beginnen die ersten Arbeiten aus dem AP6 (Transfer). Die Beispielinhalte (AP5) können ab dem 19. Monat bearbeitet werden. Sowohl AP5 als auch AP6 werden fortgesetzt bis zum Ende der Projektlaufzeit in Monat 36. Managementaufgaben (AP7) sind während der gesamten Projektlaufzeit wahrzunehmen. Für die Projektzusammenarbeit sind ein Kick-off-Workshop im 1. Monat und ein Schluss-Workshop im 36. Monat geplant, außerdem Interims-Workshops in den Monaten 7, 13, 18, 24 und 30 unter Beteiligung von allen vier Partnern. An drei der Interims-Workshops sollen Institutionen aus dem universitären und außeruniversitären Bereich teilnehmen. Für die Zusammenarbeit innerhalb der Arbeitspakete werden verteilt über die Projektlaufzeit regelmäßig Arbeitstreffen stattfinden (persönlich, Video- oder Telefonkonferenzen).

Arbeitspaket	2014												2015												2016												2017									
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	
AP1 Anforderungsanalyse																																														
AP 2 Organisationsmodell																																														
AP3 Metadaten-Strukturen																																														
AP4 Software-Entwicklung																																														
AP5 Beispielinhalte																																														
AP6 Transfer																																														
AP7 Projektmanagement																																														
Workshop			X						X							X				X								X				X					X									



## Literatur

Crosas, Mercè (2011): The Dataverse Network. An Open-Source Application for Sharing, Discovering and Preserving Data. In: D-Lib Magazine, Vol. 17, No. 1/2 doi:10.1045/january2011-crosas.

DataCite (2011): DataCite Metadata Schema for the Publication and Citation of Research Data. Version 2.2. July 2011. doi:10.5438/0005.

DESTATIS (2012): Studierende. Insgesamt nach Fächergruppen. Wintersemester 2011/2012. Online <https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/BildungForschungKultur/Hochschulen/Tabellen/StudierendeInsgesamtFaechergruppe.html> [Zugriff am 23.01.2013].

DFG (1998): Vorschläge zur Sicherung guter wissenschaftlicher Praxis: Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“. Denkschrift. Deutsche Forschungsgemeinschaft. Weinheim: Wiley-VCH. Online [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) [Zugriff am 20.2.2013].

European Union (2010): Riding the wave. How Europe can gain from the rising tide of scientific data. Final report of the High level Expert Group on Scientific Data. A submission to the European Commission. Online <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf> [Zugriff am 20.2.2013].

Feijen, Martin (2011): What researchers want. A literature study of researchers' requirements with respect to storage and access to research data. SURFfoundation. Online [http://www.surf.nl/nl/publicaties/documents/what\\_researchers\\_want.pdf](http://www.surf.nl/nl/publicaties/documents/what_researchers_want.pdf) [Zugriff am 07.02.2013].

Hausstein, Brigitte; Zenk-Möltgen, Wolfgang; Wilde, Anja; Schleinstein, Natalija (2011a): da|ra Metadatenchema Version 1.0. GESIS Working Papers 2011/14. doi:10.4232/10.mdsdoc.1.0.

Hausstein, Brigitte; Zenk-Möltgen, Wolfgang (2011b): da|ra – Ein Service der GESIS für die Zitation sozialwissenschaftlicher Daten. In: Schomburg, Silke; Leggewie, Claus; Lobin, Henning und Puschmann, Cornelius (Hrsg.): Digitale Wissenschaft. Stand und Entwicklung digital vernetzter Forschung in Deutschland. Beiträge der Tagung vom 20./21. September 2010, Köln, 2., ergänzte Fassung. S. 139–147.

Jasny, Barbara R.; Chin, Gilbert; Chong, Lisa; Vignieri, Sacha (2011): Again, and Again, and Again ... In: Science, Vol. 334, No. 6060, S. 1225, doi:10.1126/science.334.6060.1225.

Jensen, Uwe; Katsanidou, Alexia; Zenk-Möltgen, Wolfgang (2011): Metadaten und Standards. In: Büttner, Stephan; Hobohm, Hans-Christoph; Müller, Lars (Hrsg.): Handbuch Forschungsdatenmanagement, Bad Honnef: Bock u. Herchen. S. 83–100.

King, Gary (1995): Replication, Replication. In: Political Science and Politics, Vol. 28, No. 3, S. 443–499.

Knoblauch, Hubert; Solga, Heike (2011): Thesen zur Handhabung quantitativer und qualitativer Daten in Forschungsinfrastrukturen der Sozialwissenschaften: Ist eine Integration möglich? RatSWD Working Paper 190/2011. Online [http://www.ratswd.de/download/RatSWD\\_WP\\_2011/RatSWD\\_WP\\_190.pdf](http://www.ratswd.de/download/RatSWD_WP_2011/RatSWD_WP_190.pdf) [Zugriff am 20.2.2013].

Kunze, Sven R. (2012): Ein Personalisierungskonzept für Dataset-Repositorys am Beispiel von CKAN. In: Technische Universität Chemnitz – Fakultät für Informatik (Hrsg.): Studentensymposium Informatik Chemnitz 2012. Tagungsband zum 1. Studentensymposium Chemnitz vom 4. Juli 2012, Chemnitz. S.27-37.

Loeliger, Jon (2006): Collaborating With Git. A variety of clever tools allow you to share your software. In: Linux Magazine, June 2006.



McCullough, B. D. (2009): Open Access Economics Journals and the Market for Reproducible Economic Research. In: Economic Analysis & Policy, March 2009, Vol. 39 No. 1. Online [http://www.eap-journal.com.au/archive/v39\\_i1\\_9\\_mccullough.pdf](http://www.eap-journal.com.au/archive/v39_i1_9_mccullough.pdf) [Zugriff am 22.02.2013].

McMahon, Tamara M.; Powell, James E.; Hopkins, Matthew; Alcazar, Daniel A.; Miller, Laniece E.; Collins, Linn; Mane, Ketan K. (2012): Social Awareness Tools For Science Research. In: D-Lib Magazine, March/April 2012, Vol. 18, No. 3/4, doi:10.1045/march2012-mcmahon.

Open Society Institute (2004): A Guide To Institutional Repository Software. New York, 3rd edition. Online [http://www.soros.org/openaccess/pdf/OSI\\_Guide\\_to\\_IR\\_Software\\_v3.pdf](http://www.soros.org/openaccess/pdf/OSI_Guide_to_IR_Software_v3.pdf) [Zugriff am 20.2.2013].

Piwowar, Heather A.; Day, Roger S., Fridsma, Douglas B. (2007): Sharing Detailed Research Data Is Associated with Increased Citation Rate. In: PLoS ONE, Vol. 2, No. 3, e308. doi:10.1371/journal.pone.0000308.

RatSWD (Hrsg.) (2011): Auf Erfolgen aufbauend. Zur Weiterentwicklung der Forschungsinfrastruktur für die Sozial-, Verhaltens- und Wirtschaftswissenschaften. Empfehlungen des Rats für Sozial- und Wirtschaftsdaten. Opladen & Farmington Hills: Budrich.

Siegert, Olaf; Toepfer, Ralf; Vlaeminck, Sven (2012): Forschungsdatenmanagement in den Wirtschaftswissenschaften – Ausgewählte Dienste und Projekte der Deutschen Zentralbibliothek für Wirtschaftswissenschaften – Leibniz-Informationszentrum Wirtschaft (ZBW). In: Altenhöner, Reinhard; Oellers, Claudia (Hrsg.): Langzeitarchivierung von Forschungsdaten. Standards und disziplinspezifische Lösungen, Berlin: Scivero. S. 217–225.

Steinhart, Gail (2011): DataStaR: A Data Sharing and Publication Infrastructure to Support Research. In: Agricultural Information Worldwide, Vol. 4, No. 1, S. 16 –20. Online <http://journals.sfu.ca/iaald/index.php/aginfo/article/view/199> [Zugriff am 20.2.2013].

Steinhart, Gail; Chen, Eric; Arguillas, Florio; Dietrich, Dianne; Kramer, Stefan (2011): Research Data Management Service Group Survey of NSF Principal Investigators at Cornell University. Online <http://hdl.handle.net/1813/25624> [Zugriff am 20.2.2013].

Wissenschaftsrat (2012): Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020. Drs. 2359-12 vom 13.7.2012. Berlin. Online <http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf> [Zugriff am 20.2.2013].

Zaugg, Holt; West, Richard E.; Tateishi, Isaku; Randall, Daniel L. (2011): Mendeley: creating communities of scholarly inquiry through research collaboration. In: TechTrends, Vol. 55, No. 1, S. 32–36. doi: 10.1007/s11528-011-0467-y.

Zenk-Möltgen, Wolfgang; Habel, Norma (2012): Der GESIS Datenbestandskatalog und sein Metadatenschema, Version 1.8. GESIS Technical Reports, 2012/01. Online [http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis\\_reihen/gesis\\_methoden\\_berichte/2012/TechnicalReport\\_2012-01.pdf](http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/gesis_methoden_berichte/2012/TechnicalReport_2012-01.pdf) [Zugriff am 20.2.2013].