

Übersicht bewilligte Projekte aus der Ausschreibung „Informationsinfrastrukturen für Forschungsdaten“

26. April 2011

Prof. Dr. Wolfgang Wägele, Zoologisches Forschungsmuseum Alexander König Bonn; Dr. Lars Vogt, Universität Bonn; Dr. Thomas Hörnschemeyer, Universität Göttingen; Prof. Dr. Hannelore Hoch, Museum für Naturkunde, Berlin; Prof. Dr. Rolf Georg Beutel, Universität Jena

Extension and modification of Morph D Base producing a system for permanent storage and documentation of volume data of biological objects in high resolution

In den letzten Jahren hat sich die morphologisch basierte Forschung an Organismen durch die sehr schnelle Etablierung neuer Methoden, wie der Röntgentomographie (μ CT, SR μ CT), der Laser Scanning Mikroskopie (cLSM), der Digitalisierung histologischer Schnittserien, etc., stark verändert. All diese Methoden erzeugen extrem große Datenmengen, die aufgrund ihres ungewöhnlich hohen Informationsgehaltes, von besonderem wissenschaftlichem Wert sind. Bisher existiert allerdings kein Repositorium für die dauerhafte Speicherung dieser Datenmengen. Unser Projekt zielt darauf ab, Morph D Base, die zentrale Datenbank des DFG SPP 1174, für die Verwaltung großer Mengen dreidimensionaler Daten und der zugehörigen Metadaten zu optimieren. Das erweiterte Morph D Base wird vom Museum König Bonn, auch über die Laufzeit dieses Projektes hinaus, als Kommunikationsplattform für Wissenschaftler und Speichersystem für Daten zur Morphologie weiterbetrieben werden. Die Metadaten werden frei über das Internet zugänglich sein, die Dateneingabe erfordert eine Registrierung am Datenbanksystem und der Zugriff auf die Originaldaten wird der Kontrolle der Person unterliegen, die die Daten in die Datenbank eingestellt hat. Ein Standard für Metadaten zur Beschreibung von Volumendaten wird in direkter Kommunikation mit den Nutzern der Datenbank kontinuierlich angepasst und weiterentwickelt. Die wissenschaftliche Gemeinschaft wird durch Vorträge, Poster und Symposien bei den einschlägigen Tagungen über die Datenbank informiert werden.

Prof. Dr. York Sure, Mannheim, GESIS - Leibniz-Institut für Sozialwissenschaften; Christian Benz, Mannheim, Universitätsbibliothek; Prof. Dr. Heinrich Stuckenschmidt, Mannheim, Institut für Informatik

InFoLiS - Integration von Forschungsdaten und Literatur in den Sozialwissenschaften

Die wachsende Vernetzung von digitalen Informationssystemen und ihren Ressourcen spielt im digitalen Zeitalter eine wichtige Rolle und stellt ein wichtiges Erfordernis für die wissenschaftliche Informationsversorgung dar. Stehen die Ressourcen unverbunden nebeneinander, müssen die Wissenschaftlerinnen und Wissenschaftler in den empirisch ausgerichteten Sozialwissenschaften die mit den Publikationen verbundenen Forschungsdaten in den jeweiligen Volltexten suchen, in denen sie nicht standardisiert zitiert sind. Daher ist es ein wichtiges Erfordernis, dass die Metadaten der Forschungsdaten mit den bibliothekarischen Titeldaten verknüpft sind und bereits in den Recherchesystemen der Fachinformationszentren und Universitätsbibliotheken gefunden werden. Die GESIS und die Universität Mannheim möchten dieses Ziel im vorliegenden Antrag realisieren und die im Datenbestandskatalog der GESIS verwalteten Forschungsdaten des Datenarchivs mit den Aufsätzen und Büchern verknüpfen, die die Universitätsbibliothek Mannheim in einem Lucene basierten Recherchesystem präsentiert.

Die Integration von Forschungsdaten und Literatur ist im Sinne der Ausschreibung "Informationsinfrastrukturen für Forschungsdaten" eine Maßnahme, die bereits bestehende leistungsfähige Forschungsdatenrepositorien weiter professionalisiert und ausbaut. Im Rahmen des vorliegenden Antrags werden verschiedene Verfahren untersucht, technisch in den beteiligten Softwaresystemen umgesetzt und anschließend eine Empfehlung ausgesprochen, welches Verfahren am besten geeignet ist, um die oben skizzierte Integration zu realisieren und damit eine Nachnutzung durch andere Universitäten mit sozialwissenschaftlichen Disziplinen zu garantieren.

Prof. Dr. Walter G. Berendsohn, Berlin, Botanischer Garten und Botanisches Museum (BGBM); Anton Güntsch, Berlin, Botanischer Garten und Botanisches Museum, FU Berlin

Ein generisches Annotationssystem für Biodiversitätsdaten

Die Infrastruktur der Global Biodiversity Information Facility (GBIF) ermöglicht heute den globalen Zugriff auf fast 200 Millionen organismenbezogene Datensätze aus verteilten Datenbeständen mittels generischer Softwarelösungen, die z. T. in Deutschland entwickelt wurden. Deutschland steht weltweit auf dem 5. Rang unter den Datenbereitstellern und auf dem 3. Rang hinsichtlich der Abfrage von Daten aus dem GBIF-Netzwerk. In der Biodiversitätsforschung werden naturhistorische Sammlungsobjekte traditionell schriftlich annotiert, z.B. hinsichtlich ihrer taxonomischen Identität. Es gilt, einen analogen Vorgang für die im Internet verfügbaren Daten zu etablieren, also Annotationen von einzelnen Datensätzen wie auch ganzen Datenbeständen zu ermöglichen. Das beantragte Projekt soll Lösungen für eine Reihe von Domänenübergreifenden und domänenspezifischen Problemen erarbeiten und diese exemplarisch in einem Pilotsystem am Beispiel naturhistorischer Sammlungsdaten im Rahmen von GBIF-Deutschland implementieren. Zu den Problemkreisen gehören die Kategorisierung von Annotationen, die Zugangs-, Persönlichkeits- und Anerkennungsrechte von annotierenden Wissenschaftlern, die Qualitäts- bzw. Nutzbarkeitskontrolle, die Referenzierung und Verlinkung von Annotationen, die Konzeption eines die Annotation fördernden benutzerfreundlichen Systems, die Rückkoppelung zu den verteilten Datenbereitstellern, die potentielle Nutzung des System in laufenden Forschungsvorhaben zur Ausfilterung von nutzbaren Daten aus dem Gesamtsystem, und generell die Integration des Datenzugriff auf Annotationsdaten in das Gesamtsystem von GBIF, BioCASE und GBIF-Deutschland.

Prof. Dr. Rainer Schnell, Duisburg-Essen, Institut für Soziologie; Stefan Bender, Nürnberg, Institut für Arbeitsmarkt- und Berufsforschung (IAB)

Einrichtung eines Zentrums für Record-Linkage

Durch die Verknüpfung bereits vorhandener Forschungsdaten aus verschiedenen Quellen („Record-Linkage“) kann deren Analysepotential bei relativ geringem Ressourceneinsatz beträchtlich gesteigert werden. In den Human- und Sozialwissenschaften gibt es zahlreiche Anwendungsmöglichkeiten und damit Bedarf nach Record-Linkage. Eine Erschließung von Daten in diesem Sinne erfordert aber in der Regel den Einsatz technisch anspruchsvoller Record-Linkage-Techniken, da die meisten Datensätze in Deutschland nicht mit eindeutigen Identifikatoren (wie beispielsweise der Sozialversicherungsnummer) vorliegen, sondern nur über persönliche Angaben wie Name und Adresse verknüpft werden können. Eine zusätzliche Schwierigkeit erwächst aus den Datenschutzbestimmungen in Deutschland. Um die Lösung der regelmäßig anfallenden Probleme bei den einzelnen Record-Linkage-Anwendungen zu erleichtern, wird die Förderung eines überregionalen Zentrums für Record-Linkage beantragt. Das Zentrum ist als nachhaltige Infrastruktureinrichtung konzipiert, die offen für alle Fachrichtungen sein wird. Hauptziel soll es sein, durch Serviceleistungen und Forschungsarbeiten die

Zahl und Qualität von Record-Linkage-Anwendungen in den Fachwissenschaften zu steigern, und dadurch die Analysepotentiale bereits vorhandener Daten möglichst optimal zu nutzen.

Dr. Michael Diepenbroek, Universität Bremen; Dr. Michael Bittner, Deutsches Zentrum für Luft- und Raumfahrt e.V., Oberpfaffenhofen; Dr. Hannes Grobe, Alfred-Wegener-Institut für Polar- und Meeresforschung, Bremerhaven; Prof. Dr. Thomas Ludwig, Deutsches Klimarechenzentrum, Hamburg; Dr. Joachim Wächter, Deutsches GeoForschungs-Zentrum, Potsdam; TIB (Technische Informationsbibliothek und Universitätsbibliothek) Hannover

Kompetenzzentrum für Forschungsdaten aus Erde und Umwelt

Das Kompetenzzentrum ist als Bindeglied zwischen Forschungseinrichtungen, Verlagen, Bibliotheken und einem bestehenden Archivnetzwerk für Daten aus Erd- und Umweltforschung geplant. Allgemeines Ziel ist die nachhaltige Verbesserung von Datenverfügbarkeit und -qualität. Konkret sollen nachhaltige und verlässliche Wege zur Publikation wissenschaftlicher Daten geschaffen werden, die den Qualitätsstandards wissenschaftlichen Publizierens entsprechen. Dazu ist beabsichtigt, eine aus der Sicht der Nutzer einheitliche Instanz zu schaffen, welche wissenschaftlichen Projekte, Institute, Forschergruppen oder auch einzelne Wissenschaftler in allen Fragen des Datenmanagements begleitet - von der Projektplanungsphase über Datenerhebung, Qualitätssicherung, Registrierung und Langzeitarchivierung bis zur Publikation von Daten. Kernstück des Vorhabens ist die organisatorische und technische Zusammenarbeit zwischen Datenarchiven, Verlagen und Bibliotheken sowie der Aufbau einer entsprechenden webbasierten personalisierbaren Serviceplattform. Das Vorhaben ist auf die Planungen für das ICSU World Data System (WDS) abgestimmt. Teile der Organisation und Infrastruktur sollen auf das WDS übertragen werden.

Prof. Dr. Axel Börsch-Supan, Mannheim, Mannheimer Forschungsinstitut Ökonomie und Demographischer Wandel; Dr. Annelies Blom, Mannheim, Mannheimer Forschungsinstitut Ökonomie und Demographischer Wandel; Martina Brandt, Mannheim, Mannheimer Forschungsinstitut Ökonomie und Demographischer Wandel

Professionalisierung und Ausbau des Forschungsdatenzentrums Survey of Health, Ageing and Retirement in Europe (SHARE)

Ziel des beantragten Projekts in der Kategorie „Forschungsdatenrepositorien“ ist, (a) die Nutzbarkeit der SHARE Survey-Daten zu vereinfachen und zu verbessern sowie (b) Zusatzdaten zur Verfügung zu stellen. (a) SHARE ermöglicht die komparative Längsschnittanalyse komplexer Strukturen der Bevölkerungsalterung und individueller Altersprozesse in Europa. Um von der Komplexität der Daten in der Analyse zu profitieren, bedarf es an Erfahrung bei der Datenanalyse. Diese Wissenskapazität soll in der deutschen und internationalen Forschungslandschaft durch verstärkte Nutzerbetreuung, SHARE Datenanalyseurse, Austauschprogramme und die Erstellung und Verbreitung eines vereinfachten SHARE Datensatzes für die universitäre Lehre verbessert und ausgebaut werden. (b) Der geplante Ausbau der Infrastruktur des SHARE Forschungsdatenzentrums beinhaltet zudem, der Forschungsgemeinschaft zusätzlich erhobene Biomarker und verknüpfte Daten der Deutschen Rentenversicherung (Volkswagen Stiftung II/83 176) zur Verfügung zu stellen, sowie die Verfügbarkeit von Kontext-, Meta- und Paradata zu verbessern.

Prof. Dr. Dr. h.c. mult. Martin Grötschel, Berlin, Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB); Dr. Thorsten Koch, Berlin, Konrad-Zuse-Zentrum für Informationstechnik Berlin (ZIB); Prof. Dr. Ulrich Cubasch, Berlin, Institut für Meteorologie der FU Berlin; Dr. Jörn Lauterjung, Potsdam, Deutsches GeoForschungsZentrum - GFZ

Entwicklung von Workflowkomponenten für die Langzeitarchivierung von Forschungsdaten im Bereich Erd- und Umweltwissenschaften

Das Projekt ist auf die Unterstützung der Übertragung von Forschungsdaten aus unterschiedlichen Forschungsumgebungen in digitale Langzeitarchive ausgerichtet. In Anlehnung an das Konzept des „Digital Curation Continuum“ wird eine dauerhafte Domäne für die Forschungsdatenspeicherung in Form eines Langzeitarchivs entwickelt. Ein Schwerpunkt der Arbeiten besteht in der Optimierung des Ingestprozesses an der Übergabeschnittstelle zwischen (disziplinspezifischem) Datenproduzent und (disziplinunabhängigem) Langzeitarchiv. Prototypisch werden in Zusammenarbeit von Infrastruktureinrichtung und zwei unterschiedlichen Datenproduzenten aus dem Bereich Erd- und Umweltwissenschaften institutionelle Policies erstellt, Workflowkomponenten zur effizienteren Umsetzung des Ingests entwickelt und ein Langzeitarchiv als Arbeitsplattform aufgebaut. Dabei werden nach Möglichkeit existierende Softwarelösungen eingesetzt bzw. angepasst. Die Erkenntnisse fließen im Sinne der bestmöglichen Nachnutzbarkeit in Musterpolicies und Best-Practice-Dokumentationen ein. Die Nachhaltigkeit des Projekts wird durch weitgehend öffentliche Diskussion und Dokumentation des Projektfortschritts mit konkreten Ergebnissen und Nachnutzungsstrategien in der geowissenschaftlichen und fächerübergreifenden Forschergemeinschaft unterstützt. Während der gesamten Projektlaufzeit wird in einem iterativen Prozess die korrekte Interpretierbarkeit der Forschungsdaten in archivierten Informationspaketen durch Wissenschaftler und Studierende getestet.

Prof. Dr. Friederike Fless, Berlin, Deutsches Archäologisches Institut (DAI); Prof. Dr. Ortwin Dally, Berlin, Deutsches Archäologisches Institut (DAI)

Entwicklung eines Kompetenzzentrums für altertumswissenschaftliche Forschungsdaten

Die altertumswissenschaftliche Forschung steht im Bereich der Informationsinfrastrukturen von Forschungsdaten vor großen Herausforderungen. In der Vernetzung rapide anwachsender Datenbereiche, etwa von Archiv-, Photo- und Grabungsdokumentationen mit naturwissenschaftlichen und philologischen Daten, treffen Fachkulturen aufeinander, die mit Ihren Daten unterschiedlich umgehen. Die resultierenden Prozesse sind in ihren Formaten, Metabeschreibungen, Interoperabilitäts- und Archivierungsproblemen mit den derzeitigen technischen, personellen und organisatorischen Infrastrukturen nicht mehr beherrschbar: zentrale Qualitäten wie Austauschbarkeit, langfristige Verfügbarkeit und Nachnutzbarkeit werden nur in einer grundlegenden Wende erreichbar sein. Sie wäre die Grundlage für eine qualitative Steigerung von Forschungsdaten, um die immer komplexeren Fragestellungen der Altertumswissenschaften zu beantworten, eine erhöhte Sichtbarkeit von Ergebnissen im Internet zu erreichen, aber auch eine disziplinenübergreifende Verbesserung von Forschung zu befördern. Ausgehend von diesen Beobachtungen soll über das hier beantragte Projekt in einer ersten Projektphase eine substantielle Bestands- und Bedarfsanalyse durchgeführt werden, um darauf aufbauend in der zweiten Projektphase eine Organisationsstruktur zu entwickeln, die in ein Kompetenzzentrum für altertumswissenschaftliche Forschungsdaten münden soll.

Prof. Dr. Dr. h.c. mult. Ludwig M. Eichinger, Mannheim, Institut für deutsche Sprache (IDS); Dr. Andreas Witt, Mannheim, Institut für deutsche Sprache (IDS)

Zentrum für germanistische Forschungsprimärdaten

Das geplante Projekt hat die Etablierung einer einheitlichen Infrastruktur für die nachhaltige Bereitstellung von Forschungsprimärdaten aus der germanistischen Linguistik am Institut für Deutsche Sprache (IDS) zum Ziel. Damit sollen der Forschungsgemeinschaft zum einen eine empirische Basis von Primärdaten und Arbeitsmaterialien sowie eine zentrale Plattform zur Nutzbarmachung ebendieser zur Verfügung gestellt werden. Zum anderen sollen datentechnische, dokumentatorische, rechtliche Standards und Best-Practice-Richtlinien für die nachhaltige Archivierung und Verfügbarmachung von sprachwissenschaftlichen Primärdaten entwickelt werden, welche im IDS erprobt und optimiert werden und deren Anwendbarkeit für die Korpusaufbereitung in anderen Forschungsinstitutionen sichergestellt werden soll. Im Rahmen des Vorhabens soll ein einheitliches Portal zum Zwecke der Nachnutzung von Forschungsdaten entstehen, innerhalb dessen die am IDS bereits bestehenden Repositorien, die in den einzelnen Abteilungen und Bereichen verwaltet werden, nachhaltig und barrierefrei zugänglich gemacht werden. Sie umfassen die "Korpora" genannten Sammlungen von Texten sowie von Daten gesprochener Sprache und die elektronischen Lexika. Die Daten selbst verbleiben in den einzelnen Abteilungen, wo die entsprechenden Ressourcen gemäß bereichsspezifischer Expertise kontinuierlich bearbeitet und erweitert werden. In dem auf sechs Jahre angelegten Projekt sollen am IDS in einer ersten Phase zunächst die institutionellen und technischen Grundlagen für eine Infrastruktur geschaffen werden, um digitale Forschungsprimärdaten aus der germanistischen Sprachwissenschaft in einer für externe Nutzer aus dem Bereich der Wissenschaft einheitlichen Infrastruktur nachhaltig bereitstellen zu können. Hierfür soll ein Portal aufgebaut werden, das die umfangreichen Datensammlungen des IDS zusammenführt und einen zentralen Einstiegspunkt in die Datensammlungen bietet. Dazu sollen die Metadaten zugänglich gemacht und vom Portal aus auf die Ressourcen in den Repositorien der Datengeber verwiesen werden. Eine Kontrolle über den Zugriff der Daten liegt bei den Datengebern. Es soll des Weiteren die Möglichkeit geschaffen werden, einzelne Datensätze oder auch ganze Repositorien für Zwecke der Langzeitarchivierung zentral abzulegen. So können Datengeber beispielsweise Zwischenversionen oder Momentaufnahmen einer Datensammlung im Repository ablegen und so in die zentrale Langzeitarchivierung überführen. Die bei der Bearbeitung dieser Entwicklungsaufgaben anfallenden Arbeitsschritte und Entscheidungen werden auf einer kanonischen Basis, d.h. durch die Entwicklung von potenziell generalisierbaren Best-Practice-Richtlinien und Standards gelöst, die im Austausch und unter Berücksichtigung aktueller nationaler und übernationaler Entwicklungen im Bereich der Verfügbarmachung von Sprachressourcen erarbeitet werden. In einer zweiten Projektphase, d.h. ab dem vierten Jahr, sollen dann auch Repositorien und Korpora anderer Forschungseinrichtungen und externer Einzelwissenschaftler eingebunden werden. In diese zweite Entwicklungsphase fällt dann ebenfalls die Erarbeitung fortgeschrittener Konzepte, um divergierende Primärdatenformate zu erfassen, wie sie beispielsweise bei Fragebogenerhebungen oder im Zuge anderer, im IDS nicht repräsentierter linguistischer Untersuchungsmethoden entstehen. Obwohl dieses Projekt sich auf die am IDS gesammelten Ressourcen aus der germanistischen Linguistik konzentriert, weist es dennoch auch einen prototypischen Charakter auf. Dieser besteht dergestalt, dass die in den anderen Philologien am weitesten verbreiteten Datentypen, nämlich (1) linguistische Korpora geschriebener Sprache, (2) Sammlungen gesprochener Sprache, d.h. insbesondere transkribierte und ggf. annotierte Audio- und/oder Videodaten, sowie (3) elektronische Wörterbücher und lexikalische Informationssysteme, Gegenstand des Projektes sind und somit die Ergebnisse auf andere sprachwissenschaftliche Fächer übertragbar wären.

Prof. Dr. Gerold Wefer, Bremen, Zentrum für Marine Umweltwissenschaften (MARUM); Dr. Jochen Erbacher, Hannover, Bundesanstalt für Geowissenschaften und Rohstoffe (BGR); Dr. Ulrich Harms, Potsdam, Deutsches GeoForschungsZentrum - GFZ

Aufbau des Dateninformationssystems für das GESEP Kern- und Probenlager zur Erfassung und Verwaltung von Bohrkernen und Nachweis der Bestände in einem Internetportal

Die Antragsteller planen den Aufbau eines geowissenschaftlichen Proben- und Informationssystems für ein neues, nationales Lager für Bohrkern- und Proben aus Forschungsbohrungen, das initiiert durch das Deutsche Forschungsbohrkonsortium (GESEP e.V.) erstellt wird. Dabei soll am Beispiel von umfassend dokumentierten Bohrkernen und Daten des International Continental Scientific Drilling Program in einem ersten Schritt eine Probenverwaltungssoftware entwickelt und mit persistenten Identifikatoren versehen werden. Eine weitere Entwicklungslinie wird dem Aufbau eines digitalen Portals dienen, das die nachhaltige Nutzung der Daten und Proben sowie der Vernetzung mit weiteren geowissenschaftlichen Informationsinfrastrukturen ermöglicht. Diese Entwicklung soll zu einer projektunabhängigen, zentralen Datenbasis führen, die Primärdaten, Metadaten, und die dazugehörigen Proben und Publikationen zusammenführt. Dabei werden existierende Komponenten weiter entwickelt und mittels neuer Standards global verknüpft.

Prof. Dr. Frank Oliver Glöckner, Bremen, Jacobs University Bremen gGmbH School of Engineering and Science

SILVA 2.0: Building the next generation databases for ribosomal RNAs

The SILVA ribosomal RNA databases were established in 2007 to provide comprehensive, quality-checked databases of aligned rRNA (gene) sequences for all three domains of life. Since then, the SILVA user community is constantly growing and currently comprises thousands of researchers from academia and industry worldwide. In order to continue serving the community, further developments are needed to optimise the efficiency of the system for the exponentially growing number of sequences and to stabilize the code base to allow extensions requested by the users. Ongoing discussions with the user community as well as a user survey on the SILVA website indicated that improved search and retrieve functions, design and evaluation of probes and primers as well as a personalized and standardized data acquisition and submission system are needed on the website. In order to cope with the growing amount of sequence data several improvements of the SILVA pipeline like chimera detection, generation of non-redundant datasets as well as improved, automated quality checks, alignment and classification of (next generation) sequences are requested. The planned development will offer valuable services for biodiversity research, diagnostics and organism identification. It will help safeguarding primary research data by facilitating data exchange between SILVA users and the public repositories (INSDC).

Anton Güntsch, Berlin, Freie Universität Berlin Botanischer Garten und Botanisches Museum; Professor Dr. Walter Berendsohn, Berlin, Freie Universität Berlin Botanischer Garten und Botanisches Museum

Entwicklung von Workflows und Softwarekomponenten zur Rettung lebenswissenschaftlicher Primärdaten

Die Frage, wie Datenrepositorien, die derzeit in veralteten Datenbanken vorgehalten werden, langfristig für eine Nachnutzung zugänglich gemacht werden können, ist ungelöst. Eine fortwährende Migrati-

on der Systeme ist nicht möglich, da sie häufig auf unbeständigen Softwareentwicklungen aufsetzen und nicht in nachhaltigen Informationsinfrastrukturen vorgehalten werden. So gehen in immer größerem Umfang Datenbestände verloren, die mit hohem Aufwand erstellt wurden und von unschätzbarem Wert für die Wissenschaft sind. Im Projekt wird ein Verfahren entwickelt, mit dem sich Primärdaten mit vertretbarem Aufwand aus überkommenen Systemen extrahieren und in Standardformate überführen lassen. In einer ersten Phase wird mit Hilfe erweiterter existierender Datenbank-Wrapper ein Vollexport des Repositoriums in einen der einschlägigen XML-Datenstandards (z.B. ABCD, DwC, SDD, TCS, EML) generiert. Die so erzeugten Exporte sind noch keine validen XML-Dokumente, da Quellkonzepte z.T. nur in nicht standardkonformen Formaten vorliegen. Im zweiten Schritt werden sie daher mit einem zu entwickelnden Werkzeug transformiert und für die Speicherung in validen XMLFiles aufbereitet, die den Kern eines Archivs für Biodiversitätsdaten bilden, das sich durch seine Standardkonformität kontinuierlich in rezente und zukünftige Formate überführen lässt. Das Verfahren ist unabhängig von der wissenschaftlichen Domäne und damit nicht nur für die Lebenswissenschaften von Bedeutung.

Prof. Dr. Stefan Günther, Freiburg, Albert-Ludwigs-Universität Freiburg Institut für Pharmazeutische Wissenschaften

Compound Research System (CoRS): Literaturinformationssystem zur Analyse der physiologischen Wirkung von Kleinmolekülen

Die Veröffentlichung von wissenschaftlichen Erkenntnissen erfolgt weitestgehend in internationalen Journalen in digitaler Form. Diese Texte sind in der Regel unstrukturiert und bilden keine kontextbezogenen Informationen ab. Meist ist es nur Experten auf den jeweiligen Fachgebieten möglich, unter hohem Zeitaufwand zusammenhängende Informationen, die sich über die große Zahl verschiedener Artikel verteilen, in Bezug zu setzen und für eine Fragestellung richtig zu interpretieren. Die Analyse der physiologischen Wirkung von Kleinmolekülen ist eine wichtige Aufgabe im Bereich der *Life Sciences* - z.B. um potentielle neue Wirkstoffe zu identifizieren und gesundheitliche Risiken von Chemikalien abzuschätzen. Da häufig noch keine Trivialnamen für neue chemische Strukturen existieren, sind Recherchen mit zweidimensionalen Molekülbeschreibungen notwendig. Ähnliche Strukturen, deren Wirkung oder Interaktionspartner identifiziert werden konnten, liefern wichtige Informationen für die Abschätzung physiologischer Effekte. Im Rahmen des Projekts wird ein automatisiertes Molekülrecherchesystem entwickelt, das relevante Informationen aus aktueller Literatur extrahiert, Teilinformationen aus verschiedenen Quellen sinnvoll miteinander verbindet und übersichtlich präsentiert.

Prof. Dr. Wilhelm Hasselbring, Kiel, Technische Fakultät, Institut für Informatik

Publikationsprozesse für Forschungsdaten: Von der Erhebung und Verarbeitung zur Archivierung und Publikation

Ziel des Projektes PubFlow ist es, ein Konzept und eine Pilotanwendung zur Arbeit mit Forschungsdaten in Scientific Workflows zu entwickeln, um die Produktivität in der eigentlichen fachwissenschaftlichen Arbeit zu erhöhen. Dabei sollen die Prozesse von der Erhebung der Daten, der weiteren Verarbeitung bis zur Archivierung und zur Publikation betrachtet werden. Es ist nicht das Ziel, Originalbeiträge in den Bereichen der Langzeitarchivierung oder Datenzitierung zu leisten, wie es z. B. durch die PANGAEA-Datenbibliothek geleistet wird. Der Beitrag dieses Projektes besteht in der workflowbasierten Verknüpfung und Anbindung derartiger Dienste, hier aus dem Blickwinkel wissenschaftlicher Einrichtungen. In der ersten Projektphase von PubFlow soll es darum gehen, für die Wissenschaftsdiszip-

linen der Meeres- und Geowissenschaften Publikationsprozesse für Forschungsdaten als disziplinspezifische Organisationsformen zu erarbeiten und als vorbereitende Pilotanwendungen für workflowbasierte Informationsinfrastrukturen umzusetzen. Diese Pilotanwendungen sollen einerseits einen Vorbildcharakter für diese Disziplin an anderen Standorten haben. Perspektivisch sollen bereits andere Wissenschaftsbereiche wie die Lebens- und die Wirtschaftswissenschaften berücksichtigt werden, um dann in einer zweiten Projektphase weitere Wissenschaftsdisziplinen integrieren zu können.

Prof. Dr. Martin Heimann, Max-Planck-Institut für Biochemie, Jena; Prof. Dr. Birgitta König-Ries, Universität Jena; Prof. Dr. Christiane Schullius, Universität Jena; Prof. Dr. Wolfgang W. Weisser, Universität Jena

BExIS++ - Modularisierung und Skalierung der BExIS Experimentdatenhaltungsplattform für die Umweltsystemforschung

Am MPI für Biogeochemie wird seit 2007 im Rahmen des DFG SPP 1374, Biodiversitäts-Exploratorien, die Experimentdatenumgebung BExIS entwickelt. BExIS dient der Erfassung, Speicherung und Verfügbarmachung von Daten aus zurzeit 40 Teilprojekten. Es gibt großes Interesse an einer Nutzung von BExIS über das ursprüngliche Projekt hinaus. Hier bietet sich die Chance für die funktionelle Biodiversitätsforschung und darüber hinaus die Ökosystemforschung insgesamt eine weithin anerkannte, einheitliche Datenhaltungskomponente zu entwickeln. Eine solche Entwicklung geht allerdings über die im Rahmen des SPP vorgesehenen Aufgaben weit hinaus. Daher soll im Rahmen des beantragten Vorhabens BExIS um folgende Funktionalität erweitert werden: Unterstützung von unterschiedlichen Metadatenstandards und der darin kodierten Geschäftslogik, direkte Anbindung unterschiedlichster Datenproduzenten, sowie erweiterte Analysewerkzeuge. BExIS soll darüber hinaus zu einem modularisierten System, das für unterschiedliche Verbundprojekte einfach zu instanzieren und für die projektinterne - aber gruppenübergreifende - Datenhaltung genutzt werden kann, weiterentwickelt werden. Gleichberechtigt neben der technischen Entwicklung steht die Einführung einer curricularen Unterstützung, die das Thema Experimentdatenhaltung in den einschlägigen Studiengängen verankert. Eingebettet ist das Vorhaben in eine langfristige Strategie der FSU zum Aufbau eines Kompetenzzentrums Forschungsdatenmanagement und in analoge Bestrebungen des MPI BGC.

Dr. Wolfgang Ludwig, Freising, Wissenschaftszentrum Weihenstephan, Lehrstuhl für Mikrobiologie

ARB in the age of high throughput sequencing: adaption to the requirements of large scale environmental and metagenomic studies and maintenance of the respective databases

Im Rahmen des von der DFG geförderten Projektes Lu-421/5-1 wurden Teile des ARB-Softwarepaketes hinsichtlich der automatisierten Pflege integrativer Datenbanken alignierter und analysierter Nukleinsäuresequenzen und jedweder Art diesen zugeordneter Zusatzdaten erweitert und verbessert (siehe Bericht; Pruesse et al., 2008; <http://www.arb-silva.de>). Die umfangreichen Arbeiten im Zusammenhang mit der Erstellung und Pflege von qualitativ hochwertigen "Seed-" und "Core"-rRNA-Datenbanken sowie Anfragen und Anregungen aus dem Kreis der ARB- und SILVA-Nutzer haben deutlich gezeigt, dass substanzielle Erweiterungen und Verbesserungen des ARB-Paketes erforderlich sind, um den Anforderungen der Hochdurchsatzsequenzierungsära gerecht zu werden, insbesondere vor dem Hintergrund der Vielzahl an umfangreichen Umwelt- und Metagenomstudien. Neu- oder Redesign bzw. Verbesserung oder Anpassung sowie weitgehende Automatisierung von ARB-Komponenten für Datenbank-"merging", Stammbaumrekonstruktion, -optimierung und -vergleich,

Alignment von (hochvariablen) Sequenzabschnitten, "Data-mining", sowie Pflege und Analyse von Proteinsequenzdaten sind geplant. Ferner ist intensive Bereinigung des umfangreichen, in 18 Jahren unter Mitwirkung einer Vielzahl von Autoren entstandenen komplexen Codes erforderlich.

Prof. Dr. Anke Lüdeling, Berlin, Institut für deutsche Sprache und Linguistik der HU Berlin; Prof. Dr. Karin Donhauser, Berlin, Institut für deutsche Sprache und Linguistik der HU Berlin; Dr. Laurent Romary, Berlin, Institut für deutsche Sprache und Linguistik der HU Berlin; Prof. Dr. Peter Schirnbacher, Berlin, Computer- und Medienservice

Entwicklung einer nachhaltigen und nutzerorientierten Speicherung und Bereitstellung von Forschungsdaten für die historische Linguistik

Das geplante Projekt verfolgt das Ziel, für das Fachgebiet der historischen Linguistik eine Infrastruktur unter Einbeziehung eines Repository zu entwickeln, mit der fachspezifische Primärdaten in Form tief annotierter und komplex strukturierter Textkorpora erfasst, persistent gespeichert und jederzeit zugreifbar bereitgestellt werden. Textkorpora und ihre Interpretationen (Annotationen) sind die wichtigste Datenquelle für die historische Linguistik. In dem hier vorgeschlagenen System LAUDATIO sollen diese Daten (a) über ein mächtiges Suchtool online recherchiert werden, wobei die Treffermengen persistent referenzierbar sind oder (b) in standardisierten nachhaltigen Formaten in einem Repository gespeichert. Außerdem sollen die Daten über ein API für beliebige Anwendungen zur Verfügung gestellt werden können. Dies dient der Konsistenz und Transparenz wissenschaftlicher Arbeitsprozesse. LAUDATIO soll offen für historische Korpora jeglicher Herkunft sein und eine langfristig verfügbare und stabile Datenquelle für Primärdaten sowie Arbeitsinstrument für die gesamte Fachgemeinschaft der historischen Linguistik in Deutschland und international darstellen. Diese wird durch die Zusammenarbeit mit dem Computer- und Medienservice der Humboldt- Universität und die Nutzung des dort betriebenen und anzupassenden Repository verstetigt.

Prof. Dr.-Ing. Klaus-Dieter Thoben, Bremen, Maschinenbau und Verfahrenstechnik; Prof. Dr. Rainer Malaka, Bremen, Fachbereich Mathematik und Informatik; Prof. Dr.-Ing. Hans-Werner Zoch, Bremen, Stiftung Institut für Werkstofftechnik, IWT

Aufbau eines Informationssystems für werkstoffwissenschaftliche Forschungsdaten mittels Technologien zur semantischen Wissensverarbeitung

In den Werkstoffwissenschaften fallen große Mengen an Forschungsdaten durch verschiedenste Versuche an. Derzeit basieren die eingesetzten Werkzeuge des Datenmanagements hauptsächlich auf dezentralen Insellösungen, die durch einen geringen Grad an Standardisierung geprägt sind. Dies erschwert den internen, nationalen und internationalen Austausch sowie die langfristige Nutzbarkeit bzw. Verwertbarkeit der erhobenen Daten. Zusätzlich führt es zu redundanten Arbeiten und Versuchsdurchführungen und damit vermeidbaren Kosten. Ziel des Projekts ist, das bisher dezentral und heterogen organisierte Datenmanagement in eine leistungsfähige, zentral organisierte Dateninfrastruktur zu überführen, um dadurch werkstoffwissenschaftliche Forschungsdaten langfristig und nachhaltig der Forschergemeinde zur Verfügung zu stellen. Gemeinsam mit Werkstoffwissenschaftlern soll ein systematisches Vorgehen zum Umgang mit Forschungsdaten abgeleitet werden. Darauf aufbauend soll eine softwaretechnische Lösung entwickelt werden. Diese soll offen gestaltet sein, Standards unterstützen und Schnittstellen zu anderen etablierten Datenbanken und Suchmaschinen bieten.

Prof. Dr. Manfred Wischnewsky, Bremen, FB Mathematik und Informatik; Irena Medjedovic, Bremen, Bremen International Graduate School of Social Sciences (BIGSSS); Maria Elisabeth Müller, Bremen, Staats- und Universitätsbibliothek; Dr. Markus Quandt, Köln, Leibniz-Institut für Sozialwissenschaften (GESIS)

Aufbau einer Serviceeinrichtung für Daten der Qualitativen empirischen Sozialforschung. Primärdaten für eScience in den Sozialwissenschaften

Ziel des Projektes ist es, organisatorische und technische Strukturen zur Speicherung, Referenzierung und Verfügbarkeit von qualitativen Forschungsprimärdaten in den Sozialwissenschaften zu initiieren, konzeptuell zu gestalten und bedarfsgerecht zu entwickeln. Es werden grundlegende Funktionseinheiten des Forschungsdatenservice implementiert und deren Anwendung modellhaft anhand qualitätsgeprüfter Studien erprobt. Dabei wird auf internationale Standards wie das Metadatenschema der Data Documentation Initiative und den Erfahrungen des ESDS Qualidata, das international als Vorreiter im Bereich Datenservice qualitativer Forschungsdaten gilt, aufgebaut. Dienste werden in Form von WSDLVSoap-Webservices im Rahmen eines Workflow Referenzmodells der Workflow Management Coalition bereitgestellt. In Zusammenarbeit mit dem Leibniz-Institut GESIS, dessen Service auf quantitative Daten ausgerichtet ist, wird durch den Aufbau der Serviceeinrichtung deutschlandweit erstmals der Zugang zu qualitativen Primärdaten ermöglicht. Um den zentralen Nachweis aller sozialwissenschaftlichen Primärdaten bei GESIS zu unterstützen, werden Verfahren zur Integration der Metadaten der archivierten Studien in das von GESIS geplante umfassende Portal entwickelt. Die Vernetzung mit weiteren qualitativen Datenbeständen erweitert den Nutzerkreis.

Gerald Neumann, Berlin, Berlin-Brandenburgische Akademie der Wissenschaften; Dr. Stephan Wiederkehr, Berlin, Berlin-Brandenburgische Akademie der Wissenschaften

Wissensspeicher – Daten geisteswissenschaftlicher Grundlagenforschung

Die Berlin-Brandenburgische Akademie der Wissenschaften (BBAW) ist die größte außeruniversitäre geisteswissenschaftliche Forschungseinrichtung in der Region. An ihr wird insbesondere geisteswissenschaftliche Grundlagenforschung betrieben, die an Universitäten nicht realisiert werden kann. Die Ergebnisse vieler Akademienvorhaben sind, neben den Print-Publikationen, zunehmend elektronisch aufbereitete Daten der Grundlagenforschung, die in dieser Form einmalig sind und mit großem Aufwand und in hoher Qualität gesammelt wurden. Die Organisation des einfachen Zugriffs auf diese Daten stellt für die BBAW eine große Herausforderung dar. Seit 2003 bündelt die BBAW ihre elektronischen Aktivitäten in der Telota-Initiative (Telota = The electronic life of the academy). Aufgabe von Telota ist es, die Akademienvorhaben beim Ausbau ihres elektronischen Workflows zu unterstützen und die elektronische Präsentation der Forschungsergebnisse umzusetzen. Derzeitiger Schwerpunkt der Bemühungen von Telota zur Schaffung einer langfristigen Infrastruktur für die Forschungsdaten sind die Arbeiten am digitalen Wissensspeicher der BBAW. Der digitale Wissensspeicher soll die vorhandenen elektronischen Ressourcen der BBAW erfassen und bündeln. Er stellt eine Weiterentwicklung der BBAW-Infrastruktur für die Präsentation der digitalen Wissensbestände dar und wird die Sichtbarkeit der Materialien und die Nutzungsmöglichkeit der Daten durch die Forschung im Internet erhöhen. Grundlage eines erfolgreichen Konzepts ist die Integration des Wissensspeichers in die bestehende Infrastruktur. Er soll sich von Entwicklungen aus dem Bereich der Wissensorganisation und -Präsentation im virtuellen Raum inspirieren lassen und diese im Sinne der besonderen Situation an der Akademie weiterentwickeln. Ziel des beantragten Projekts ist es, die nächste Ausbaustufe des bestehenden Wissensspeichers aufzubauen. Damit fällt der Antrag unseres Erachtens in die Förderkategorie „(b) Forschungsdatenrepositorien“, da ein bereits bestehendes und etabliertes System ausgebaut und um leistungsfähige Funktionen erweitert werden soll. Die Erschließung der Ressourcen über Metadaten soll verfeinert und Möglichkeiten der Verknüpfung von Metadaten geschaffen werden. Außerdem werden die Retrievalmöglichkeiten um eine linguistische Volltextsuche erweitert, die auch Sprachen wie das Altgriechische, Latein und Arabisch abdecken wird. Konzeption und Implementie-

rung einer dialoggestützten Retrievaloberfläche sollen den Zugriff auf die Forschungsdaten unterstützen und steuern. Die Einrichtung von Schnittstellen und Services wird auch externen Anbietern den Zugriff auf die Metadaten und die Verarbeitung der Ressourcen ermöglichen. Verbunden mit dem Ausbau der standardisierten Schnittstellen für eine automatische Weiterverarbeitung der Daten wird das beantragte Projekt die strukturierte inhaltliche Bereitstellung aller Forschungsdaten der Akademie auf ein neues Niveau heben.

Dr. Markus Quandt, Köln, GESIS - Leibniz-Institut für Sozialwissenschaften; Prof. Dr. York Sure, Bonn, GESIS - Leibniz-Institut für Sozialwissenschaften

da|ra - Aufbau einer Registrierungsagentur für sozialwissenschaftliche Forschungsdaten

Ziel des Vorhabens unter dem Titel da|ra ist der Aufbau eines Nachweis- und Registrierungssystems für Forschungsdaten der Sozialforschung im deutschsprachigen Raum. Darüber sollten Datenbestände, d.h. ihre Titel, Themen, Provenienz und Zugangsmöglichkeiten so umfassend wie möglich nachgewiesen werden. Damit würde erstmals das Ziel erreichbar, Transparenz über die Datenverfügbarkeit herzustellen. In Zusammenarbeit mit qualifizierten, datenhaltenden Institutionen soll darüber hinaus eine geeignete Teilmenge der Bestände mit so genannten Persistent Identifiern (PIDs) versehen werden. PIDs gewährleisten die eindeutige Zitierbarkeit von Forschungsdaten und stiften somit Anreize für die Datenveröffentlichung. Zudem erlauben sie die Referenzierung in technischen Systemen, so dass z.B. die Vernetzung von Daten und Publikationen vereinfacht wird. Beide Nachweisstufen werden prinzipiell unabhängig vom physischen Speicherort der Forschungsdatensätze anwendbar sein. Das System soll nach Abschluss des Projektes ein offener Service sein, in dem Forscher Datenbestände auch selbständig nachweisen und qualifizierte Registrierungspartner selbständig PIDs registrieren können.

Prof. Dr.-Ing. Klaus Rheidt, Universität Cottbus; Prof. Dr.-Ing. Ingo Schmitt, Universität Cottbus; Prof. Dr.-Ing. Frank Schwarzbach, Hochschule für Technik und Wirtschaft, Dresden

OpenInfRA - Ein webbasiertes Informationssystem zur Dokumentation und Publikation archäologischer Forschungsprojekte

Bei altertumswissenschaftlichen Feldforschungen unter Beteiligung unterschiedlicher Disziplinen werden heute Daten zumeist in digitaler Form gesammelt. Da diese häufig nicht nachhaltig gesichert, veröffentlicht und vernetzt sind, soll ein webbasiertes Informationssystem entwickelt und für die archäologischen Fächer und die Bauforschung frei zur Verfügung gestellt werden, sowohl für Projekte an Universitäten, Museen, Akademien, dem DAI und anderen Institutionen als auch für ausländische Kooperationspartner. OpenInfRA wird ein breites Spektrum von Arbeitsmethoden und Themen berücksichtigen, individuell zu modifizieren und auf strukturelle Einheitlichkeit ausgerichtet sein, um eine projektübergreifende Auswertung von Daten zu vereinfachen. Durch die Bereitstellung umfangreicher Recherchefunktionen sowie die Einbindung externer Ressourcen und Applikationen wird ein hohes Maß an Interoperabilität angestrebt. Neue Funktionen zur Verarbeitung, Analyse und Präsentation von 3D-Geometrien werden entwickelt, um räumliche Zusammenhänge durch die Integration webbasierter GIS-Funktionalitäten besser darstellen zu können. Die Möglichkeit zur zeitnahen, digitalen Publikation umfassender Datenbestände und deren Kommentierung durch Nutzer soll die zeit- und ortsunabhängige Zusammenarbeit von Wissenschaftlern sowie neue Formen des Diskurses in der Altersforschung fördern.

Prof. Dr. Otto Rienhoff, Göttingen, Universitätsmedizin Göttingen, Abteilung Medizinische Informatik

Langzeitarchivierung biomedizinischer Forschungsdaten

Ziel des Vorhabens ist es, präzise Anforderungen an die Langzeitarchivierung für Bildverarbeitung und Genomforschung in der Biomedizin abzuleiten. Wegen der Komplexität der Materie soll dies mit einer mehrstufigen Methodik in Zusammenarbeit mit der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) erfolgen. Die Präzisierung der Anforderungen der beiden Teilbereiche wird mit Hilfe von Use Cases vorgenommen. Auf Grundlage der Anforderungen aus den Teilbereichen sollen übergreifende Handlungsempfehlungen in Form eines Betriebskonzepts für die biomedizinische Community abgeleitet werden. Dieses Betriebskonzept spezifiziert die organisatorische Einbindung und technische Implementierung. Dabei werden weitere wissenschaftliche Disziplinen und Institutionen zum Erfahrungsaustausch eingebunden. Am Ende des Projekts soll eine begutachtungsfähige Laborimplementierung am Standort Göttingen vorliegen. Die Laborimplementierung soll in einer möglichen Anschlussförderung weit in die Community getragen werden.

Dr. Thomas Schmidt, Sonderforschungsbereich 538, Universität, Hamburg

Etablierung eines Schwerpunkts "Mehrsprachigkeit und gesprochene Sprache" im Hamburger Zentrum für Sprachkorpora

Das Projekt zielt auf den Ausbau, die weitere Professionalisierung und die Verstetigung der Arbeit des Projekts Z2 „Computergestützte Erfassungs- und Analysemethoden multilingualer Daten“, das von der DFG im Rahmen des SFB 538 „Mehrsprachigkeit“ an der Universität Hamburg seit Juli 2005 bis Juni 2011 gefördert wird. Gegenstand der Arbeit sind computergestützte Methoden der Erstellung, Verwaltung, Auswertung und Publikation von Korpora gesprochener Sprache, also von systematischen Zusammenstellungen von Audiooder Videoaufnahmen natürlich sprachlicher Interaktionen und deren Transkriptionen. Ergebnis des hier beantragten Projektes soll ein ausgearbeitetes Konzept sein, wie die im Zuge der EXMARaLDA-Entwicklung aufgebaute Expertise, das Software-System selbst und die damit erstellten Datenbestände über das Hamburger Zentrum für Sprachkorpora und im Rahmen entstehender Forschungsinfrastrukturen der wissenschaftlichen Gemeinschaft dauerhaft zur Verfügung gestellt werden können. Zu diesem Zwecke wird erstens ein Datenrepositorium konzipiert und implementiert, über das existierende Datenbestände verwaltet, gepflegt und für eine Weiter- oder Wiederverwendung bereitgestellt werden. Zweitens werden verschiedene Instrumente für einen professionellen Anwendersupport entwickelt und erprobt.

Ekkehart Seusing, Kiel, Leibniz Informationszentrum Wirtschaft, (ZBW); Prof. Dr. Gert G. Wagner, Berlin, Deutsches Institut für Wirtschaftsforschung (DIW); Prof. Dr. Joachim Wagner, Lüneburg, Fakultät Wirtschaft, Institut für Volkswirtschaftslehre

EDaWaX (European Data Watch Extended)

In der empirischen Wirtschaftsforschung gibt es derzeit kaum effektive Möglichkeiten, Forschungsdaten im Kontext der zugehörigen Textpublikationen zur Förderung der Replikation, Kritik und des wissenschaftlichen Diskurses bereit zu stellen bzw. zu nutzen. Ziel von EDaWaX ist es, erstmals im Rahmen eines ganzheitlichen Ansatzes ein publikationsbezogenes Datenarchiv zu entwickeln, welches sich den damit verbunden Herausforderungen annimmt. Die von EDaWaX angebotene Lösung ist deshalb ganzheitlich, weil sie zunächst auf Basis einer fachwissenschaftlich fundierten Analyse die

heutigen Schwierigkeiten, wie insbesondere fehlende Anreizstrukturen bei der Replikation von Forschungsergebnissen und der Mehrfachnutzung von Daten systematisch aufbereitet. Parallel dazu werden bereits existierende Lösungen im Kontext von Datenarchiven und wirtschaftswissenschaftlichen Fachzeitschriften sowie die rechtlichen Rahmenbedingungen im Hinblick auf die Eignung für die Anforderungen aus EDaWaX untersucht. Auf Basis der Analyseergebnisse wird in einem zweiten Schritt ein Metadatenschema für die Beschreibung und Auszeichnung der Daten entwickelt und ein Regelwerk zur schemakonformen Transformation der Beschreibungen und Daten erstellt. Die Erkenntnisse münden schließlich im dritten Schritt in die Pilotanwendung eines innovativen publikationsbezogenen Datenarchivs am Beispiel der renommierten Fachzeitschrift „*Schmollers Jahrbuch / Journal of Applied Social Science Studies*“. Ein Projekt mit der Komplexität von EDaWaX kann nur dann erfolgreich umgesetzt werden, wenn komplementäre Kompetenzen über die Projektpartner eingebracht werden. Diese Voraussetzungen einschließlich hervorragender Netzwerke, die speziell für die Analyse sowie eine nachhaltige Verwertung der Projektergebnisse bedeutsam sind, bringen die beiden Antragsteller RatSWD und ZBW ein.

Dr. Dirk Walther, Potsdam-Golm, Max-Planck-Institut für molekulare Pflanzenphysiologie (MPIMP); Dr. Joachim Kopka, Potsdam, Max-Planck-Institut für molekulare Pflanzenphysiologie (MPIMP)

Development of the Golm Metabolome Database as a central plant metabolomics information resource

Metabolomics has emerged as a central technology to study cellular processes. In recent years, standardized methodological and computational approaches have been developed rendering metabolomics datasets obtained for different samples and by different research groups comparable and creating new opportunities for their integrative analysis. This potential is currently not utilized in the metabolomics field and important data are typically dispersed as supplementary files. Adequate centralized information infrastructure resources are still largely missing. We plan to establish the information infrastructure that allows researchers to deposit, query as well as analyze metabolomic profile data. We will build upon the already developed Golm Metabolome Database (GMD) that is focused on plant metabolomic data. We propose to further develop the GMD to integrate metabolite profile datasets and to create the infrastructure specifically tailored towards their web-based presentation, integration and analysis. Furthermore, we plan to integrate and upload comprehensive plant metabolite profiling datasets to serve as reference sets (Plant Metabolite Atlas (PMA)) for other researchers. Thus, we intend to establish a central and broadly available plant-centric metabolomics profile data resource and analysis platform.

Prof. Dr. Christof Wolf, Mannheim, Leibniz-Institut für Sozialwissenschaften - GESIS; Prof. Dr. York Sure, Bonn, Leibniz-Institut für Sozialwissenschaften - GESIS

Forschungsbasierte Metadaten für amtliche Erhebungen: Ausbau von MISSY

Gegenstand des Projektantrags ist die inhaltliche und technologische Erweiterung des Mikrodaten-Informationssystems MISSY. MISSY wurde teils mit finanzieller Förderung des BMBF bei GESIS entwickelt und unterstützt die sozialwissenschaftliche Profession bei Forschungsarbeiten mit dem Mikrozensus. Zukünftig sollen in MISSY weitere amtliche Mikrodaten strukturiert dokumentiert werden. Hier wird die Integration der Metadaten zu drei amtlichen Erhebungen beantragt, die von der Scientific Community stark genutzt werden: die Einkommens- und Verbrauchsstichprobe (EVS), die Europäi-

sche Arbeitskräftestichprobe (EU-LFS) und die europäische Statistik zu Einkommen und Lebensbedingungen (EU-SILC). Neben der inhaltlichen Aufbereitung und Eingabe der Metadaten ist eine Erweiterung der technologischen Basis notwendig. Das Erfassungswerkzeug und die Zugänge zum Metadatenangebot im Web müssen erweitert werden. Diese Erweiterungen betreffen vor allem die Mehrsprachigkeit und die Dokumentation auf Länderebene im Falle der europäischen Daten. Zur Sicherstellung der Nutzung der Projektergebnisse wird zum einen die Nachnutzung der Metadaten durch andere Anbieter (z. B. Forschungsdatenzentren) unterstützt. Zum anderen werden die zentralen Dienste von MISSY als Open-Source-Software bereitgestellt.