

Siegert, Olaf; Toepfer, Ralf

Working Paper

Von der Sammlung gedruckter Statistiken zum Datenrepository. Wie sich das Forschungsdatenmanagement in der ZBW verändert und welche Rolle die Wissenschaft dabei spielt

RatSWD Working Paper, No. 172

Provided in Cooperation with:

German Data Forum (RatSWD)

Suggested Citation: Siegert, Olaf; Toepfer, Ralf (2011) : Von der Sammlung gedruckter Statistiken zum Datenrepository. Wie sich das Forschungsdatenmanagement in der ZBW verändert und welche Rolle die Wissenschaft dabei spielt, RatSWD Working Paper, No. 172, Rat für Sozial- und Wirtschaftsdaten (RatSWD), Berlin

This Version is available at:

<https://hdl.handle.net/10419/75324>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

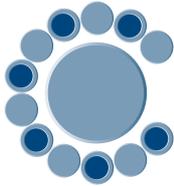
Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Rat für Sozial- und
Wirtschaftsdaten (RatSWD)

www.ratswd.de

RatSWD

Working Paper Series

Working Paper

Nr. 172

Von der Sammlung gedruckter Statistiken zum Datenrepository

Wie sich das Forschungsdatenmanagement in der ZBW
verändert und welche Rolle die Wissenschaft dabei spielt

Olaf Siegert und Ralf Toepfer

Februar 2011

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Working Paper Series des Rates für Sozial- und Wirtschaftsdaten (RatSWD)

Die *RatSWD Working Papers* Reihe startete Ende 2007. Seit 2009 werden in dieser Publikationsreihe nur noch konzeptionelle und historische Arbeiten, die sich mit der Gestaltung der statistischen Infrastruktur und der Forschungsinfrastruktur in den Sozial-, Verhaltens- und Wirtschaftswissenschaften beschäftigen, publiziert. Dies sind insbesondere Papiere zur Gestaltung der Amtlichen Statistik, der Ressortforschung und der akademisch getragenen Forschungsinfrastruktur sowie Beiträge, die Arbeit des RatSWD selbst betreffend. Auch Papiere, die sich auf die oben genannten Bereiche außerhalb Deutschlands und auf supranationale Aspekte beziehen, sind besonders willkommen.

RatSWD Working Papers sind nicht-exklusiv, d. h. einer Veröffentlichung an anderen Orten steht nichts im Wege. Alle Arbeiten können und sollen auch in fachlich, institutionell und örtlich spezialisierten Reihen erscheinen. Die *RatSWD Working Papers* können nicht über den Buchhandel, sondern nur online über den RatSWD bezogen werden.

Um nicht deutsch sprechenden Nutzer/innen die Arbeit mit der neuen Reihe zu erleichtern, sind auf den englischen Internetseiten der *RatSWD Working Papers* nur die englischsprachigen Papers zu finden, auf den deutschen Seiten werden alle Nummern der Reihe chronologisch geordnet aufgelistet.

Einige ursprünglich in der *RatSWD Working Papers* Reihe erschienen empirischen Forschungsarbeiten, sind ab 2009 in der RatSWD Research Notes Reihe zu finden.

Die Inhalte der *RatSWD Working Papers* stellen ausdrücklich die Meinung der jeweiligen Autor/innen dar und nicht die des RatSWD.

Herausgeber der RatSWD Working Paper Series:

Vorsitzender des RatSWD (2007/2008 Heike Solga; seit 2009 Gert G. Wagner)

Geschäftsführer des RatSWD (Denis Huschka)

Von der Sammlung gedruckter Statistiken zum Datenrepository
**Wie sich das Forschungsdatenmanagement in der ZBW verändert und
welche Rolle die Wissenschaft dabei spielt***

Olaf Siegert und Ralf Toepfer

*ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften
Leibniz-Informationszentrum Wirtschaft*

Abstract:

Forschungsdaten waren schon immer ein Thema für Bibliotheken, allerdings bislang eher in einem sehr eingeschränkten Sinne, nämlich vor allem dort, wo es sich um gedruckte Publikationen oder Lizenzdatenbanken handelt: Diese wurden und werden gesammelt bzw. über Nachweissysteme zur Verfügung gestellt.

Mit der zunehmenden Digitalisierung und Vernetzung nicht nur von Formaten sondern auch von Arbeitsprozessen in der Forschung ergeben sich für Bibliotheken neue Möglichkeiten des Wissenschaftssupports. Jetzt können sie die Forschende nicht nur als Endnutzer von Fachinformation versorgen, sondern auch Dienste für den übrigen Forschungs- und Publikationsprozess anbieten. Ein Arbeitsfeld dabei ist der Bereich Forschungsdaten.

In der ZBW spielen diese eine zunehmend größere Rolle bei Forschungs- und Innovationsprojekten. Dabei reicht das Spektrum von Aktivitäten im Kontext von Open-Access-Projekten, bei denen Forschungsdaten als „related data“ einbezogen werden über die Aufbereitung von historischen Statistiken für die Weiterverarbeitung mit Tabellenkalkulationsprogrammen bis hin zu Datenarchiven für Forschernetzwerke.

Das Paper gibt einen Überblick über die verschiedenen Projektvorhaben mit Forschungsdatenbezug in der ZBW. Außerdem wird beleuchtet, wie die Wirtschaftsforschung jeweils in die Aktivitäten einbezogen ist.

* Dieses Working Paper ist die Ausarbeitung eines gleichnamigen Vortrags auf dem Deutschen Bibliothekartag 2011 in Berlin

Einleitung

Forschungsdatenmanagement und Bibliotheken – Wie passt das zusammen? Hier muss man zunächst unterscheiden zwischen den klassischen Bibliotheksaufgaben der Sammlung und Aufbereitung von Publikationen und neueren Arbeitsbereichen der letzten Jahre, bei denen sich Bibliotheken in einem erweiterten Arbeitsauftrag auch um Infrastrukturaufgaben kümmern, die klassischerweise bei Verlagen oder Rechenzentren angesiedelt waren. Im Bereich ihrer klassischen Aufgaben haben sich Bibliotheken dort um Forschungsdaten gekümmert, wo diese als (Verlags-) Veröffentlichung daherkamen. Bestes Beispiel dafür sind die Serien des statistischen Bundesamtes, die von vielen Bibliotheken bezogen und über ihre Kataloge nachgewiesen wurden.

Im Zeitalter des Internets verschwindet diese Publikationsform zunehmend. An ihre Stelle treten neue Erscheinungsformen von Forschungsdaten, die auf eine sich wandelnde Bibliothekswelt treffen. In seinen übergreifenden Empfehlungen zu Informationsinfrastrukturen weist der Wissenschaftsrat zu recht darauf hin, dass neben den traditionellen Aufgaben wissenschaftlicher Bibliotheken „...neue Funktionen wie der Betrieb von potenziell global nutzbaren Repositorien zur Sicherung eines langfristigen Zugangs zu digitalen Medien oder Forschungsdaten [treten].“¹ Auch der Rat für Sozial- und Wirtschaftsdaten (RatSWD) hält fest, dass in den aktuellen Diskussionen zur Neuausrichtung der Informationsinfrastruktur wissenschaftliche Bibliotheken im allgemeinen und zentrale Fachbibliotheken im besonderen eine wichtige Rolle bei der Dokumentation sowie der Bereitstellung von Forschungsdaten spielen könnten.²

Vor diesem Hintergrund, soll im Folgenden am Beispiel verschiedener Projekte und Aktivitäten in der ZBW – Deutschen Zentralbibliothek für Wirtschaftswissenschaften gezeigt werden, welche Ausprägungen das Forschungsdatenmanagement in Bibliotheken annehmen kann. Dabei lassen sich die Anwendungen im Wesentlichen in drei Bereiche unterteilen:

- Forschungsdaten als Textpublikationen
- Forschungsdaten im Kontext von elektronischem Publizieren

1 Wissenschaftsrat: Übergreifende Empfehlungen zu Informationsinfrastrukturen. Drs. 10466-11, Berlin 28.01.2011, S. 14; <http://www.wissenschaftsrat.de/download/archiv/10466-11.pdf>

2 Vgl. German Data Forum (RatSWD) (Hg.): Building on Progress: Expanding the Research Infrastructure for the Social, Economic, and Behavioral Sciences. Vol. 1, Budrich UniPress Ltd., Opladen & Farrington Hills, MI 2010, p. 39. <http://www.budrich-unipress.de/media/products/0611760001296569161.pdf>

- Forschungsdaten für spezifische Fachcommunities

1 Forschungsdaten als Textpublikationen

Traditionellerweise haben wissenschaftliche Bibliotheken Forschungsdaten meist in Form von veröffentlichten amtlichen Statistiken (z.B. Reihen des Statistischen Bundesamts etc.) gesammelt. Diese Statistiken wurden wie sonstige Publikationen als Serien oder Monographien katalogisiert und in die Bibliothekskataloge und bibliographischen Datenbanken aufgenommen. Diese Herangehensweise verwundert nicht, da die Bibliotheken seit jeher auf das Management von Publikationen fokussiert sind. Die Wissenschaft spielt dabei als Stakeholder meist nur eine indirekte Rolle (z.B. in Bibliothekskommissionen von Universitäten). Sie ist vor allem als Konsument der Dienstleistung vorgesehen, jedoch zumeist nicht in Detail an der Art der Aufbereitung einbezogen.

Eine neue Variante der klassischen Bibliotheksaufgaben stellen Digitalisierungsaktivitäten dar. Dabei werden die gedruckten Sammlungen durch die Digitalisierung für den elektronischen Zugriff über das Web einfacher zugänglich gemacht – interessierte Nutzer müssen die Werke nicht mehr bestellen oder vor Ort ausleihen, sondern können sie in der Regel frei zugänglich im Internet ansteuern.

Auf Basis von Digitalisierungen lassen sich aber auch amtliche Statistiken leichter nutzen. Ein Beispiel dafür stellt das Projekt „Digitale Reichsstatistik“³ dar. Dabei wird bislang nur in gedruckter Form vorliegendes statistisches Material aus dem Bibliotheksbestand der ZBW digitalisiert und nach den Prinzipien des Open Access frei im Internet angeboten. Die amtlichen Statistiken sind dabei sowohl als Imagedatei im PDF-Format als auch in einer für Tabellenkalkulationsprogramme verarbeitbaren Form nutzbar. Das hierfür notwendige Verfahren wird exemplarisch anhand der sogenannten „Alten Folge“ der „Statistik des Deutschen Reichs“ (Zeitraum 1873-1883) entwickelt und in der Praxis erprobt. Abbildung 1 verdeutlicht das geplante Verfahren. Aufbauend auf der Imagedigitalisierung werden die in der „Reichsstatistik“ enthaltenen Dateien doppelt manuell erfasst und nach dem Dokumentenformat TEI⁴ ausgezeichnet. Anschließend erfolgt eine Transformation in das gewünschte Zielformat (z.B. csv, HTML, Stata, SPSS etc.)

3 http://www.zbw.eu/ueber_uns/projekte/reichsstatistik.htm

4 TEI: Text Encoding Initiative, siehe: <http://www.tei-c.org/index.xml>

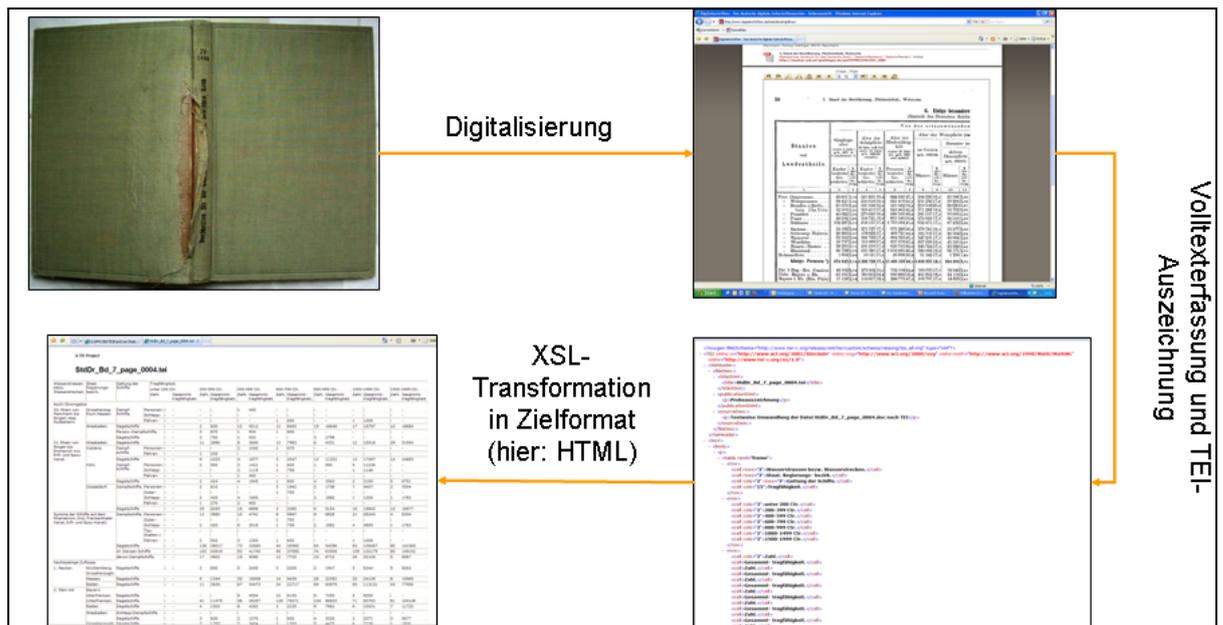


Abbildung 1: Workflowschema der digitalen Aufbereitung der Statistik des Deutschen Reichs

Durch die digitale Aufbereitung wird die Quellennutzung wesentlich vereinfacht. So kann unter anderem in den Inhaltsverzeichnissen online recherchiert werden und die Inhalte werden zusätzlich über ein thematisch geordnetes Register erschlossen. Darüber hinaus werden aus den Imagedateien Volltextdigitalisate erstellt, um daraus herunterladbare Dateien zu generieren sowie eine Volltextsuche anbieten zu können. Ziel des Projekts ist es, dass jede in der „Reichsstatistik“ enthaltene Tabelle von den Nutzern komfortabel heruntergeladen und für eigene Berechnungen genutzt werden kann. Mit diesem Projekt schlägt die ZBW die Brücke von der klassischen Bibliotheksaufgabe zur zeitgemäßen Aufbereitung von Bibliotheksbeständen für die modernen Nutzungsanforderungen der empirischen Forschung.

2 Forschungsdaten im Kontext von elektronischem Publizieren (textzentriertes Modell)

Einen anderen Zugang zum Thema Forschungsdaten bietet der Bereich des elektronischen Publizierens. Hier unterstützen Bibliotheken Autoren und Herausgeber von Veröffentlichungen durch die Bereitstellung geeigneter Archivstrukturen oder persistenter Adressierung zur Bereitstellung von For-

schungsdaten als Ergänzung zu Textpublikationen. Die ZBW ist hier gleich mit mehreren Projekten aktiv.

a) *daIra*

daIra⁵ (Datenregistrierungsagentur) ist ein Service für Datenproduzenten und datenhaltende Organisationen zur Registrierung von Forschungsdaten – der Fokus liegt dabei derzeit auf sozialwissenschaftlichen Daten. Die Idee von daIra ist, dass im Web zugänglichen Forschungsdatensätze dauerhafte stabile Internetadressen (sog. Persistent Identifier) erhalten. Dies ermöglicht eine leichtere Zitierbarkeit der Datensätze und damit eine höhere eigene Sichtbarkeit, z.B. in Forschungspublikationen. Den gleichen Weg sind vor einigen Jahren die großen Fachzeitschriften gegangen, die für Ihre Online-Ausgaben ebenfalls Persistent Identifier in Form von DOIs (Digital Object Identifier) verwenden.

daIra richtet sich vor allem an Datenarchive, Forschungsdatenzentren, Service- datenzentren und zukünftig auch an individuelle Forscher. Diese können damit z.B. Umfragedaten, Aggregatdaten, Mikrodaten oder Daten aus Quellenstudien registrieren und mit DOIs versehen lassen. Der Service wurde von GESIS aufgebaut und wird zukünftig im Kontext von DataCite gemeinsam mit der ZBW betrieben, um Datenproduzenten im Bereich Wirtschafts- und Sozialwissenschaften einen Service aus einer Hand bieten zu können⁶. Die Forschungscommunity wird in Form eines wissenschaftlichen Beirats eingebunden, der die strategische Weiterentwicklung des Service begleitet.

b) *Economists Online*

Ein Anwendungsbeispiel für die Integration eines Datenarchivs in ein Publikationsportal stellt Economists Online⁷ dar. Dabei wurde im Rahmen eines EU-Projekts ein Open-Access-Portal aufgebaut, das im Wesentlichen auf den institutionellen Repositorien von 20 europäischen Universitäten und Forschungseinrichtungen basiert und die kompletten Publikationslisten von etwa 1000 Wirtschaftsforschern nachweist - nach Möglichkeit inklusive Volltext im Open Access, was bei etwa 40% der Veröffentlichungen der Fall ist (ca. 35.000

5 <http://www.thesis.org/dara>

6 http://www.zbw.eu/presse/pressemitteilungen/2010_12_01.htm

7 <http://www.economistsonline.org/>

Volltexte). Die Autoren wurden zudem gebeten, bei empirischen Arbeiten ihre verwendeten Daten, Programme und Dokumentationen an die Repositorien zu liefern, damit diese zur Replikation und für Re-Analysen ebenfalls im Open Access zur Verfügung stehen. Die gelieferten Datensätze werden in einem Datenarchiv⁸ inklusive Metadatenbeschreibung und Persistent Identifier bereitgestellt (vgl. Abbildung 2).

The screenshot shows the 'ECONOMISTSONLINE' website. The main content area displays the title 'THE DETERMINANTS OF ALLOWANCE PRICES IN THE EUROPEAN EMISSIONS TRADING SCHEME: CAN WE EXPECT AN EFFICIENT ALLOWANCE MARKET 2008? [DATASET]' with the handle 'hdl:1902.1/13854'. Below the title are tabs for 'Cataloging Information', 'Documentation, Data and Analysis', 'User Comments', and 'Versions'. A prominent blue box contains a citation instruction: 'If you use these data, please add the following citation to your scholarly references. Why cite?' followed by a text box with the citation: 'W. Rickels; V. Duscha; A. Keller; S. Peterson, 2007, "The determinants of allowance prices in the European emissions trading scheme: Can we expect an efficient allowance market 2008? [Dataset]", http://hdl.handle.net/1902.1/13854 Dataverse Network [Distributor] V2 [Version]'. Below this is a 'Citation Format' dropdown menu with a 'Print' button. The 'Data Citation Details' section lists: Study Global ID (hdl:1902.1/13854), Authors (W. Rickels, V. Duscha, A. Keller, S. Peterson), Producer (Kiel Institute for the World Economy), Production Date (2007), Distributor (Dataverse Network (DVN)), Deposit Date (November 09, 2009), and Provenance (Network of European Economists Online Dataverse). The 'Abstract and Scope' section shows Keywords (Emissions permit (STW), Price (STW)) and Topic Classification (C22 (JEL); Q58 (JEL); Q56 (JEL); Q54 (JEL)).

Abbildung 2: Nachweis eines Datensatzes zu einer Publikation in Economists Online

Allerdings sind nur sehr wenige Forscher der Bitte nachgekommen (insgesamt etwa 100 Datensätze). Die Hintergründe wurden im Rahmen einer projektbezogenen Nutzerstudie⁹ untersucht. Dabei wurden folgende Gründe für die Zurückhaltung bei der Herausgabe von Daten identifiziert:

- Bedenken vor Nutzung durch konkurrierende Forscher/innen
- Sorge, dass Urheberschaft nicht anerkannt wird

⁸ <http://dvn.iq.harvard.edu/dvn/dv/NEEO>

⁹ http://www.neeoproject.eu/NEEO_UserStudy_1.pdf

- Konflikt mit Verlagen (ausschließliche Nutzungsrechte)
- rechtliche Bedenken, wenn Basisdaten aus lizenzierten Datenbanken entnommen wurden
- generelle Vorbehalte gegenüber Open Access von Forschungsdaten

Die wissenschaftliche Begleitung des Projekts erfolgte durch ein Scientific Advisory Board¹⁰, das aus renommierten europäischen Wirtschaftsforschern bestand.

c) *Open-Access-Journal „Economics“*

Zusammen mit dem Institut für Weltwirtschaft hat die ZBW das Open-Access-Journal „Economics“ aufgebaut, das seit 2007 online ist. Da unter den Einreichungen auch viele empirische Beiträge sind, stellte sich die Frage, wie man mit den Berechnungen und Rohdaten umgeht bzw. wie man sie in den Publikationsprozess einbezieht. Ähnlich wie mehrere renommierte US-Fachzeitschriften hat man sich bei Economics dafür entschieden, eine sogenannte „Data Availability Policy“ zu verabschieden. Diese besagt, dass Autoren empirischer Papers nach Akzeptanz ihres Beitrags alle notwendigen Daten, Programme und Beschreibungen/Dokumentationen an das Journal schicken, die für eine Replikation der Berechnungen nötig sind¹¹. Ergänzend dazu wurde für das Journal ein Datenarchiv¹² aufgebaut, bei dem die Quellen mit bibliographischen Metadaten beschrieben werden (vgl. Abbildung 3).

Außerdem wird für jeden Datensatz ein Persistent Identifier (in Form eines sog. „handle“) vergeben, um eine eigenständige Zitation unabhängig vom Zeitschriftenaufsatz zu ermöglichen. Die inhaltliche Steuerung des gesamten Journalprozesses inklusive Datenarchiv erfolgt durch den Herausgeber Institut für Weltwirtschaft, während die ZBW als Publikationsdienstleister ihr bibliothekarisches Know-how einbringt.

10 <http://www.nereus4economics.info/scientific-advisory-board.html>

11 <http://www.economics-ejournal.org/submission/data-availability-policy>

12 <http://www.economics-ejournal.org/special-areas/data-sets>

The screenshot displays the Economics Dataverse interface. At the top left is the 'Economics' logo with the tagline 'The Open-Access, Open-Assessment E-Journal'. To the right, it says 'POWERED BY THE Dataverse Network PROJECT'. Below the header, there are navigation links: 'Search', 'User Guides', 'Report Issue', 'Log In', and 'Create Account'. The main content area shows the title 'SHOULD WE TRUST THE EMPIRICAL EVIDENCE FROM PRESENT VALUE MODELS OF THE CURRENT ACCOUNT? [DATASET]' and the DOI 'hdl:1902.1/13900'. Below the title, there is a navigation bar with tabs for 'Cataloging Information', 'Documentation, Data and Analysis' (which is active), 'User Comments', and 'Versions'. A message box explains the download process: 'Use the check boxes next to the file name to download multiple files. Data files will be downloaded in their default format. You can also download all the files in a category by checking the box next to the category name. You will be prompted to save a single archive file. Study files that have restricted access will not be downloaded.' Below this message is a 'Download All Selected Files' button and a list of files with checkboxes and download links. The files listed are: 'Annual.csv' (Plain Text - 22 KB - 9 downloads), 'Estimation 3 lags.sas' (SAS Syntax - 6 KB - 9 downloads), 'FTests.sas' (SAS Syntax - 2 KB - 9 downloads), 'Quarterly.csv' (Plain Text - 75 KB - 9 downloads), 'simulationsvardgpbelgium.m' (Plain Text - 6 KB - 9 downloads), 'simulationsvardgpcanada.m' (Plain Text - 8 KB - 8 downloads), 'simulationsvardgpdk.m' (Plain Text - 6 KB - 9 downloads), 'simulationsvardggsueden.m' (Plain Text - 6 KB - 11 downloads), and 'simulationsvardgpk.m' (Plain Text - 6 KB - 9 downloads).

Abbildung 3: Einzelelemente eines Datensatzes im Datenarchiv von „Economics“

d) *EdaWax (European Datawatch Extended)*

Das DFG-geförderte Projekt EdaWax hat das Ziel, im Rahmen eines ganzheitlichen Ansatzes ein publikationsbezogenes Datenarchiv für die Fachzeitschrift „*Schmollers Jahrbuch / Journal of Applied Social Science Studies*“¹³ zu entwickeln.

Um dies zu erreichen, ist eine umfassende Analyse der bestehenden Lösungen und Rahmenbedingungen notwendig. Dies erfolgt in der ersten Analysephase des Projekts. Dabei werden zunächst auf Basis einer fachwissenschaftlich fundierten Analyse die Anreizprobleme untersucht, die bislang verhindern, dass Daten für Replikationsanalysen in adäquater Form bereitgestellt und genutzt werden. Parallel dazu werden bereits existierende Lösungen im Kontext von Datenarchiven und wirtschaftswissenschaftlichen Fachzeitschriften sowie die rechtlichen Rah-

13 <http://schmollersjahrbuch.diw.de/schmollersjahrbuch/>

menbedingungen 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*“. Das auf 30 Monate angesetzte Projekt wird von der ZBW in Kooperation mit dem RatSWD¹⁴ und dem Institut für Innovationsforschung, Technologiemanagement und Entrepreneurship an der LMU München¹⁵ durchgeführt.

Als Ergebnis des Projektes wird ein publikationsbezogenes Datenarchiv implementiert sein, dass die veröffentlichten Textpublikationen und die dafür verwendeten Forschungsdaten sowie ergänzenden Dokumentationen in einem zusammenhängenden Kontext präsentiert und damit für Dritte nachvollziehbar macht. Mit EdaWaX wird der wissenschaftlichen Community erstmals eine technologische Umgebung angeboten werden, die auf konkreten Nutzerbedürfnissen fußt und die völlig neuartige Formen der empirischen Wirtschaftsforschung ermöglichen kann.

3 Forschungsdaten für spezifische Fachcommunities (community-zentriertes Modell)

Im Gegensatz zum publikationsbezogenen Modell, bei dem die Daten jeweils als Informationsergänzung zur Textpublikation verstanden werden, geht es beim community-zentrierten Modell um die Aufbereitung von Forschungsdaten innerhalb einer Forschungscommunity und unabhängig von Veröffentlichungen (wobei Publikationen sehr wohl auf Basis dieser Daten entstehen können). Zudem steht im community-zentrierten Modell nicht die freie Verfügbarkeit von Daten im Vordergrund, sondern vielmehr das Datenmanagement und der Datenaustausch innerhalb eines geschützten Bereichs, von z.B. verteilt

¹⁴ <http://www.ratswd.de/>

¹⁵ <http://www.inno-tec.bwl.uni-muenchen.de/personen/professoren/harhoff/index.html>

arbeitenden Forschergruppen. Dies erfolgt nach dem Prinzip einer Virtuellen Forschungsumgebung, d.h. es geht um Werkzeuge und Technologien, die von den Forschern benötigt werden, um kollaborativ, international und institutionell unabhängig interagieren zu können.¹⁶

a) *REDECO (Research Data in Economics)*

Im Rahmen des geplanten EU-Projekts REDECO arbeiten 16 europäische Universitäten, Bibliotheken und Datenzentren (darunter aus Deutschland die Universität Konstanz, GESIS und die ZBW) zusammen, um eine Dateninfrastruktur für verteilt arbeitende Forschergruppen aufzubauen. Dabei geht es pilothaft zunächst um die wirtschaftswissenschaftlichen Subdisziplinen Finanzmarktökonomie, Ökonometrie, Experimentalökonomie, und Wirtschaftsgeschichte. Im Rahmen des Projekts ist vorgesehen, dass Forschende dieser Bereiche zusammen mit den Infrastrukturanbietern Anforderungen für eine virtuelle Forschungsumgebung erarbeiten, die zum Management und zum Teilen von Forschungsdaten geeignet ist. In einem weiteren Arbeitspaket ist dann die technische Umsetzung dieser Anforderungen vorgesehen. Die erarbeitete Lösung soll in Workshops mit den Forschenden vorgestellt und weiter verbessert werden. Im Rahmen von REDECO sollen auch die Bereiche Rechtsfragen (Urheberrecht und Datenschutz) und Datenbeschreibung (Metadaten) behandelt und werden und die Eignung von Forschungsdaten für Linked Open Data untersucht werden.

b) *MaWiFo*

MaWiFo (Management wirtschaftswissenschaftlicher Forschungsdaten) ist ein Kooperationsprojekt der ZBW mit der Universität Kiel¹⁷. Es verfolgt einen ähnlichen Ansatz wie REDECO, nämlich den Aufbau einer Forschungsdateninfrastruktur im Sinne einer Virtuellen Forschungsumgebung. Zielgruppe dieses Projekts sind die Forschenden der wirtschaftswissenschaftlichen Fakultät der Universität.

Im Rahmen des Projekts werden nicht nur die vorliegenden Forschungspriärdaten aufbereitet und archiviert, sondern darüber hinaus auch Anreizsysteme für die Forschenden geschaffen, damit diese ihre Daten öffentlich bereitstellen.

¹⁶ Vgl.: http://www.jisc.ac.uk/programme_vre.html

¹⁷ http://www.zbw.eu/presse/pressemitteilungen/2010_09_21.htm

Dies kann nur über einen Weg gelingen, bei dem die Datenaufbereitung als eigenständige wissenschaftliche Leistung anerkannt wird¹⁸. Auf der technischen Seite liegt der Fokus auf dem Einsatz semantischer Technologien, um die Forschungsprimärdaten mit Veröffentlichungen in einen Bedeutungszusammenhang zu bringen.

An MaWiFo beteiligt ist neben der ZBW auch das Institut für Informatik der Universität Kiel. Während das Institut für Informatik seine IT-Kompetenz einbringt, trägt die ZBW neben ihrem technischen Know-how vor allem mit inhaltlichen Aspekten und Anforderungen ihrer Kundengruppen aus den Wirtschaftswissenschaften zum Erfolg bei. Im Rahmen der Exzellenzinitiative der Universität werden für das Projekt mehrere Dissertationsstellen eingerichtet.

4 Fazit

Für wissenschaftliche Bibliotheken reicht es nicht mehr aus, der Scientific Community zukünftig nur klassische Services wie den (lokalen) Zugang zu Büchern und Fachzeitschriften anzubieten. Vielmehr ist es notwendig, die Wissenschaft als Infrastrukturanbieter im gesamten Publikations- und Forschungsprozess zu unterstützen. Hier besteht in einer zunehmend digitalen und vernetzten Forschungslandschaft auch ein erhöhter Bedarf auf Seiten der Forschenden – wie dies erst kürzlich in den Empfehlungen des Wissenschaftsrats zu den Informationsinfrastrukturen konstatiert wurde¹⁹.

In diesem Kontext gewinnt das Forschungsdatenmanagement zunehmend an Bedeutung. Die Rolle der wissenschaftlichen Bibliotheken im Zusammenhang mit Forschungsdaten ist sicherlich noch nicht hinreichend ausdiskutiert, kann jedoch vielfältige Formen annehmen. Wie die dargelegten Anwendungsbeispiele der ZBW zeigen, können Informationsinfrastruktureinrichtungen aufgrund ihrer hohen Metadatenkompetenz, ihrer großen Erfahrung im Bereich der Archivierung sowie nicht zuletzt aufgrund der Dauerhaftigkeit ihrer Aufgaben eine wichtige Rolle einnehmen. Klar ist aber auch, dass diese neuen Aufgaben nur gemeinsam

18 Im Editorial der Zeitschrift Nature wurde hierfür die einprägsame Formulierung „Credit where credit is due“ postuliert (Nature 462, 825 (17 December 2009) | doi:10.1038/462825a, die auch vom RatSWD (vgl. German Data Forum (RatSWD), a.a.O., S. 38) aufgegriffen wurde.

19 <http://www.wissenschaftsrat.de/download/archiv/10466-11.pdf>

mit den Stakeholdern, insbesondere der wissenschaftlichen Forschung, erfolgreich gestaltet werden kann.