

Zur Rolle von Forschungsdatenzentren beim Management von publikationsbezogenen Forschungsdaten – Ergebnisse einer Befragung von wissenschaftlichen Infrastrukturdienstleistern im Bereich der Sozial- und Wirtschaftswissenschaften

Im vorliegenden Beitrag werden die Ergebnisse einer Analyse zusammengefasst, in der untersucht wurde, ob, und wenn ja welche Services für das Management von publikationsbezogenen Forschungsdaten gegenwärtig bei wissenschaftlichen Infrastrukturdienstleistern in den Sozial- und Wirtschaftswissenschaften bestehen. Die Analyse wurde Mithilfe von Desktop-Research sowie einer Online-Befragung, an der sich 22 von 46 angeschriebenen Institutionen beteiligten, durchgeführt. Untersucht wurden vor allem deutsche und europäische Forschungsdatenzentren, Bibliotheken und Archive. Insbesondere wurde untersucht, ob diese Organisationen extern erzeugte Forschungsdaten, den dazugehörigen Berechnungscode (Syntax) und ggf. genutzte (selbstgeschriebene) Software grundsätzlich speichern und hosten. Weitere Themenfelder waren Metadatenstandards, Persistente Identifikatoren, Verfügbarkeit von Schnittstellen (APIs) und Unterstützung von semantischen Technologien.

This paper summarizes the findings of an analysis of scientific infrastructure service providers (mainly from Germany but also from other European countries). These service providers have been evaluated with regard to their potential services for the management of publication-related research data in the field of social sciences, including economics. For this purpose we conducted a desk research as well as an online survey among 46 research data centres, library networks and public archives; almost 22 institutions responded to our survey. Topics of our survey covered storage and hosting of externally generated research data, of the code of computation (syntax) and (self-compiled) software components that have been used for statistical analyses. Other topics comprised metadata standards, persistent identifiers, availability of interfaces (APIs) and support of semantic web technologies.

EINLEITUNG

Obwohl zunehmend mehr Publikationen in sozial- und wirtschaftswissenschaftlichen Fachzeitschriften auf Basis der Auswertung von empirischen Daten erscheinen, gibt es in diesen empirischen Wissenschaften jedoch bislang kaum einfache zu handhabende und effektive Möglichkeiten, die publizierten Forschungsergebnisse im Kontext der zugehörigen Artikel zu replizieren und zu prüfen. Insbesondere wird auch die »Nachnutzung« von Daten zur Unterstützung des wissenschaftlichen Diskurses kaum unterstützt. Dies ist gerade vom wissenschaftlichen Standpunkt aus gesehen unbefriedigend, da zentrale Anforderungen an die Überprüfbarkeit empirischer Forschung, wie sie etwa im von Karl R. Popper beschriebenen Prinzip des »Kritischen Rationalismus« dargelegt werden, nicht oder nur teilweise erfüllt werden. Es fällt zudem auf, dass

die einer Publikation zugrunde liegenden Daten, selbst dann, wenn sie auf allgemein zugänglichen Datensätzen (z.B. ALLBUS¹ oder SOEP²) beruhen, in der Regel nicht in ihrer spezifischen Auswahl und Bereinigung (als finales Arbeits-File) archiviert werden. Damit sind Replikationen zwar nicht generell ausgeschlossen, aber eine Replikation fällt bei anspruchsvollen Analysen schwer.

Die derzeitige Situation stellt sowohl die Fachdisziplinen als auch wissenschaftliche Infrastrukturdienstleister wie Bibliotheken und Datenzentren vor Herausforderungen, die nur durch das Zusammenwirken der verschiedenen Akteure (z.B. Fachgesellschaften, Infrastrukturanbieter) in den Fachdisziplinen gelöst werden können.

Auf der Ebene der benötigten Infrastruktur könnten insbesondere die in den letzten zehn Jahren verstärkt etablierten Forschungsdatenzentren (FDZs) eine wichtige Rolle für das Management solcher publikationsbezogenen Daten einnehmen, da einige dieser Institutionen (Datenarchive) teilweise bereits seit Jahrzehnten über Erfahrungen im Umgang mit disziplinären Daten verfügen. Speziell Forschungsdatenzentren aus dem Bereich der Sozial- und Wirtschaftswissenschaften stehen daher im Fokus dieses Artikels, da diese Institutionen aufgrund ihres spezifischen Know-hows einen relevanten Beitrag für Speicherung, Hosting und Langzeitarchivierung solcher Forschungsdaten leisten können.

Der Beitrag beleuchtet einleitend zunächst Gründe dafür, weshalb publizierte sozial- und wirtschaftswissenschaftliche Forschungsergebnisse oftmals nicht replizierbar sind. Da sich diese Gründe teilweise auch auf der Ebene der wissenschaftlichen Infrastruktur verorten lassen, werden in der Folge die Ergebnisse einer Untersuchung der Services von Forschungsdatenzentren aus den Sozial- und Wirtschaftswissenschaften in Bezug auf das Management von publikationsbezogenen Forschungsdaten (also von Daten, die Grundlage wissenschaftlicher Publikationen in den beiden Fachdisziplinen sind) vorgestellt. Abschließend wer-



Sven Vlaeminck



Gert G. Wagner

publizierte Forschungsergebnisse lassen sich kaum replizieren

den Schlüsse für die weitere Entwicklung einer Anwendung gezogen, die das Management von publikationsbezogenen Forschungsdaten für Fachzeitschriften aus den Sozial- und Wirtschaftswissenschaften erleichtern.³

Weshalb ist sozial- und wirtschaftswissenschaftliche Forschung oft nicht replizierbar?

Die Gründe für die fehlende Replizierbarkeit sozial- und wirtschaftswissenschaftlicher Forschung lassen sich auf verschiedenen Ebenen verorten. Insbesondere die beiden erstgenannten Aspekte sind dabei jedoch nicht spezifisch für die Sozial- und Wirtschaftsforschung, sondern haben weitgehend für alle wissenschaftlichen Disziplinen Gültigkeit:

- Zum ersten sind fehlende Anreizmechanismen für Forschende zu nennen, die von ihnen bearbeiteten Forschungsdaten mit anderen Wissenschaftler/innen zu teilen (Data Sharing). Das Wissenschaftssystem belohnt die Arbeit des Aufbereitens und der Dokumentation von Daten für die Forschung, die überdies oft sehr zeitintensiv ist, nicht. Dies steht im drastischen Gegensatz zur Wertigkeit von klassischen Publikationen, die, wie Anderson et al. feststellen, nur »[...] the advertising for the data and code that produced the published results« sind (Anderson et al., 2008, S. 101).
- Forschende fürchten zudem, dass Data Sharing für sie mit Nachteilen verbunden sein könnte: Denn die bereits unter großem Arbeitsaufwand aufbereiteten Daten werden auch einer wissenschaftlichen Community zur Verfügung gestellt, die sie nutzen kann, obwohl von deren Seite kein Beitrag zur Datenaufbereitung geleistet wurde. Dies kann zu einer Schieflage in der Forschung führen, da Wissenschaftler/innen, die neue Daten generieren und für andere aufbereiten, dafür keine Reputation erlangen und es somit schwerer haben, Universitätskarrieren zu verfolgen. Zudem befürchten viele Forscher/innen eine Art Missbrauch der Daten durch Dritte, zum Beispiel durch falsche Interpretation oder durch Nutzung der Daten ohne korrekte Zitation der Urheberin bzw. des Urhebers. Schließlich ist die Rechtslage bei der Weitergabe von Datensätzen in vielen Fällen nicht ausreichend geklärt, was ebenfalls zu einer großen Zurückhaltung im Bereich »Data Sharing« führt (Siegert, Toepfer und Vlaeminck, 2012).⁴
- Derzeit verfügen nur wenige sozial- und wirtschaftswissenschaftliche Fachzeitschriften über Richtlinien, die den Umgang mit den verwendeten Forschungsdaten im Detail regeln (McCullough, 2009): Sogenannte Data Availability Policies bei-

spielsweise verpflichten Autor/inn/en empirischer Artikel dazu, die zur Erlangung der Ergebnisse verwendeten Forschungsdaten sowie die erstellte Syntax der statistischen Auswertung gemeinsam mit dem publizierten Artikel bereitzustellen. Solche Richtlinien folgen somit oftmals dem vom Gary King formulierten »Replication Standard« (King, 1995).

- Es werden nur selten geeignete Infrastrukturkomponenten für das Management von publikationsbezogenen Forschungsdaten eingesetzt, so dass oftmals nicht einmal uniforme Zitationen der hinterlegten Daten möglich sind. Lösungen, wie beispielsweise Dataverse⁵, ein leistungsfähiges Instrument für das Management und die Dokumentation von publikationsbezogenen Forschungsdaten, werden nur von wenigen Fachzeitschriften eingesetzt. Fraglich ist in diesem Kontext unter anderem, wie professionelle Forschungsdatenzentren mit solchen Daten verfahren, ob (und wenn ja welche) Services angeboten werden.

kaum Anreize zum Data Sharing

Das DFG-geförderte Forschungsprojekt EDaWaX (European Data Watch Extended – www.edawax.de) untersucht seit Herbst 2011 unter anderem die oben genannten Aspekte (vgl. für eine Übersicht der bisherigen Projektergebnisse: Vlaeminck, Wagner, Wagner, Harhoff und Siegert, 2013). Verschiedene Publikationen aus dem Projektkontext befassen sich zudem mit Teilaspekten des Projekts: So analysiert ein Artikel das Data Sharing Verhalten von Wirtschaftsforschenden aus dem Bereich der angewandte Volkswirtschaftslehre (Andreoli-Versbach und Mueller-Langer, 2013). Andere Publikationen befassten sich mit der Analyse des Forschungsdatenmanagements in wirtschaftswissenschaftlichen Fachzeitschriften (Vlaeminck, 2013; Vlaeminck und Siegert, 2012).

Ergänzend zu diesen Publikationen beschreibt dieser Aufsatz die Ergebnisse einer Evaluierung von wissenschaftlichen Infrastrukturdienstleistern in Bezug auf deren Services für das Management von publikationsbezogenen Forschungsdaten in den Sozial- und Wirtschaftswissenschaften.

geeignete Infrastrukturkomponenten werden selten eingesetzt

Bieten sozial- und wirtschaftswissenschaftliche Forschungsdatenzentren Services für publikationsbezogene Forschungsdaten-Archivierung an?

Speziell Forschungsdatenzentren könnten eigentlich ideale Institutionen für das Management von Forschungsdaten sein, die als Anhänge zu publizierten Artikeln bei Fachzeitschriften eingereicht werden. Diese Kompetenz rührt aus teils jahrzehntelangen Erfahrungen im Umgang mit den verschiedensten Arten so-

wenige Fachzeitschriften verfügen über Richtlinien für Forschungsdaten

Forschungsdatenzentren sind mehr als klassische Daten-Archive

zial- und wirtschaftswissenschaftlicher Daten, der Metadatenkompetenz bei der Erfassung und Auszeichnung von Studien und nicht zuletzt den Erfahrungen in Bezug auf den Zugriff auf diese Daten (vgl. Research Information Network, 2011). Bei den Forschungsdatenzentren handelt es sich nicht nur um »klassische Daten-Archive« (wie z.B. das Datenarchiv für Sozialwissenschaften in Köln), sondern inzwischen überwiegend um Zentren, die spezifische Daten und damit verbundene Services für die Re-Analyse ihrer Daten anbieten (zu nennen ist etwa das FDZ des Sozio-oekonomischen Panels am DIW Berlin; vgl. z.B. Huschka et al., 2012). Ein diskriminierungsfreier Datenzugang und der damit verbundene Service werden vom Rat für Sozial- und Wirtschaftsdaten (RatSWD)⁶ geprüft und sind Teil der Akkreditierungsbestimmungen für Forschungsdatenzentren.

Im Rahmen des EDaWaX-Projekts wurde daher eine Studie durchgeführt, die evaluierte, ob, und wenn ja welche Services für solche publikationsbezogenen Forschungsdaten bei Infrastrukturdienstleistern wie Forschungsdatenzentren (FDZs), Bibliotheken oder auch Archiven existieren. Zunächst wurde eine Liste von 46 Organisationen aus dem Bereich der sozial- und wirtschaftswissenschaftlichen Infrastrukturdienstleister erstellt. Darunter fanden sich sowohl die vom Rat für Sozial- und Wirtschaftsdaten (RatSWD) akkreditierten Forschungsdaten- und Datenservicezentren, die im CESSDA-Verbund⁷ organisierten zu meist europäischen Datenzentren, Bibliotheksverbünde sowie einzelne Bibliotheken und Archive aus der Bundesrepublik.

All diese Organisationen wurden zunächst per Internet-Recherche auf mögliche Serviceangebote für

Speicherung und Hosting von publikationsbezogenen Forschungsdaten untersucht. Diese Online-Recherche ergab, dass am ICPSR (Inter-university Consortium for Political and Social Research – University of Michigan) ein Publication Related Archive (PRA)⁸ vorliegt, was bereits von zahlreichen Autor/inn/en für die Hinterlegung von publikationsbezogenen Forschungsdaten genutzt wurde.⁹ Das niederländische DANS EASY¹⁰ bietet zwar keine dezidierten Services für publikationsbezogenen Forschungsdaten an, jedoch ist es prinzipiell auch möglich, DANS EASY für die Hinterlegung solcher Daten zu nutzen.¹¹

Darüber hinaus ergab die reine Desktop-Recherche jedoch wenige Anhaltspunkte für weitergehende Analysen, weshalb genauere Auskünfte durch eine detailliertere Onlinebefragung bei diesen wissenschaftlichen Infrastrukturdienstleistern eingeholt werden sollten.

DIE BEFRAGUNG

Ein Onlinefragebogen wurde im Oktober 2012 an insgesamt 46 Organisationen verschickt – darunter 35 nationale und internationale Forschungsdatenzentren (FDZs) und Datenservicezentren (DSZs), ein Archiv, sieben Bibliotheksverbünde und Bibliotheken sowie drei weitere Institutionen. 22 Organisationen beteiligten sich an der Befragung (48 %). Die Rücklaufquote ist – gemessen an den Rücklaufquoten schriftlicher Erhebungen – als sehr gut anzusehen. Bedingt durch die Struktur des Fragebogens beantworteten nicht alle Institutionen jede der gestellten Fragen. Abweichungen der Anzahl der Beantwortungen sind u.a. dadurch zu erklären.

Wichtiger als die Rücklaufquote insgesamt ist freilich die Struktur der Antwortenden bzw. Nicht-Antwortenden. Dabei zeigt sich, dass die allermeisten Antworten (86 %) aus den deutschen FDZs und DSZs, sowie den Datenzentren des CESSDA-Verbundes kamen. Deutlich unterrepräsentiert sind in den Antworten die deutschen Bibliotheksverbünde und Archive, aber auch die drei unter »Sonstiges« berücksichtigten außer-europäischen Forschungsdatenzentren. Für die Bibliotheksverbünde und das angeschriebene Archiv kann nur gemutmaßt werden, dass keine entsprechenden Services oder Angebote zum Bereich Forschungsdatenmanagement bestehen und daher eine Beantwortung des Fragebogens aus diesem Grunde nicht erfolgte. Alle folgenden Angaben beziehen sich auf die 22 Einrichtungen, die geantwortet haben.

INHALTLICHE AUSWERTUNG

Zunächst wurde mit der Befragung untersucht, ob die antwortenden Institutionen publikationsbezogene

Services von 46 Infrastrukturdienstleistungen wurden untersucht

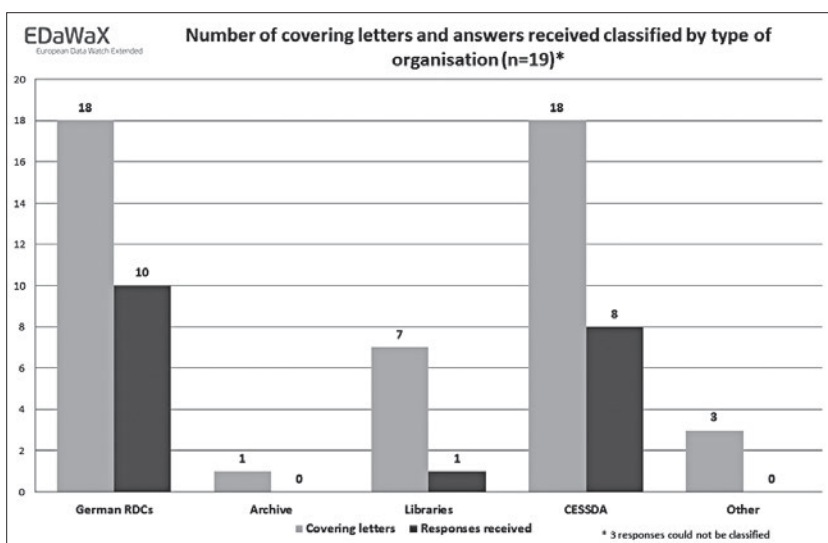


Abb. 1: Anzahl der versendeten Anschreiben / der teilnehmende Organisationen an der Befragung nach Art der Einrichtung

Forschungsdaten prinzipiell speichern und hosten.¹² Zudem wurde ermittelt, ob auch selbstgeschriebene Software und der Berechnungscode von statistischen Auswertungen durch die Institutionen gespeichert und bereitgestellt werden. All diese Daten sind oftmals Bestandteil der Einreichungen bei wirtschaftswissenschaftlichen Fachzeitschriften mit entsprechenden Policies.¹³

Datensätze

Von den untersuchten Organisationen akzeptierten mehr als drei Viertel externe Datensätze zur Speicherung. Der größte Anteil der Befragten gab dabei an, solche Forschungsdaten nur anzunehmen, wenn gewisse Kriterien erfüllt sind. Solche Kriterien bestehen etwa in Form der spezifischen thematischen Zuständigkeiten der deutschen FDZs, aber auch in Form von regionalen, überregionalen und/oder fachspezifischen Zuständigkeiten. Zudem wurden technisch-organisatorische Aspekte (Dokumentation, Maschinenlesbarkeit) und rechtliche Fragestellungen als Kriterien genannt. Von den antwortenden Organisationen gaben etwa 74 % an, solche Forschungsdaten auch zu hosten. Wenn dafür Kriterien bestehen, wurde erneut vor allem die fachliche Ausrichtung der Institution als Kriterium genannt.

Software

In Bezug auf Speicherung und Hosting von (selbstgeschriebener) Software, wie sie etwa im Rahmen wirtschaftswissenschaftlicher Simulationen Verwendung findet, zeigte sich, dass nur eine Minderheit von knapp einem Viertel der antwortenden Organisationen eine Speicherung und das Hosting solcher Daten ohne Einschränkungen unterstützt. Weitere 17 % betonten, dass auch für solche Software Kriterien existieren, wonach entschieden wird, ob Speicherung und Hosting erfolgen (beispielsweise, wenn diese wichtig für die Datenanalyse sind). Einige Organisationen gaben an, dass entsprechende Services für die Zukunft geplant seien oder solche Software als Teil der Dokumentation zu Datensätzen bereitgestellt wird. Somit sind die Speicherung und das Hosting solcher Software als »Gaps« anzusehen, die bislang nur eine übersichtliche Zahl an Organisationen anbietet.

Syntax

Fast 70 % der antwortenden Organisationen bieten die Möglichkeit, Syntax zu speichern und bereitzustellen – ein Viertel der untersuchten Organisationen tut dies jedoch nicht und plant dies auch nicht für die Zukunft. Ein Befragter nannte zudem das Kriterium, dass dies nur bei *derived variables* sinnvoll sei.

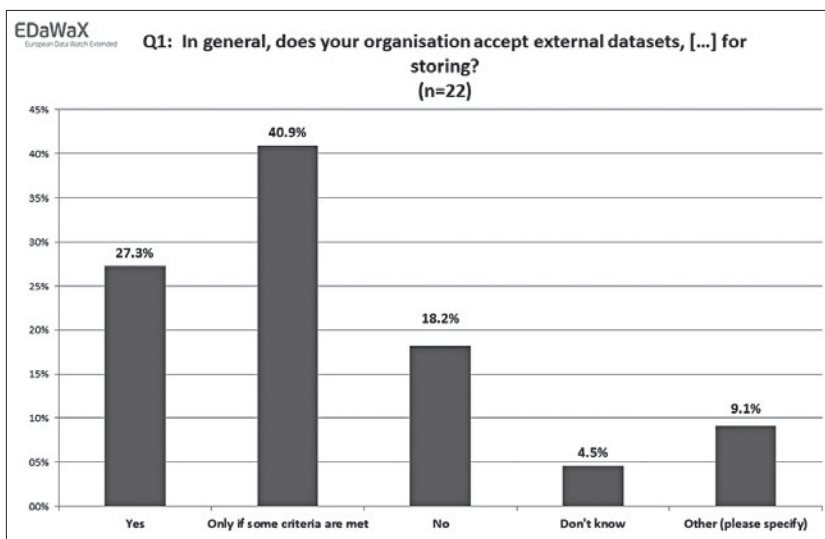


Abb. 2: Bereitschaft zur Speicherung von extern erzeugten Forschungsdatensätzen an den befragten Einrichtungen

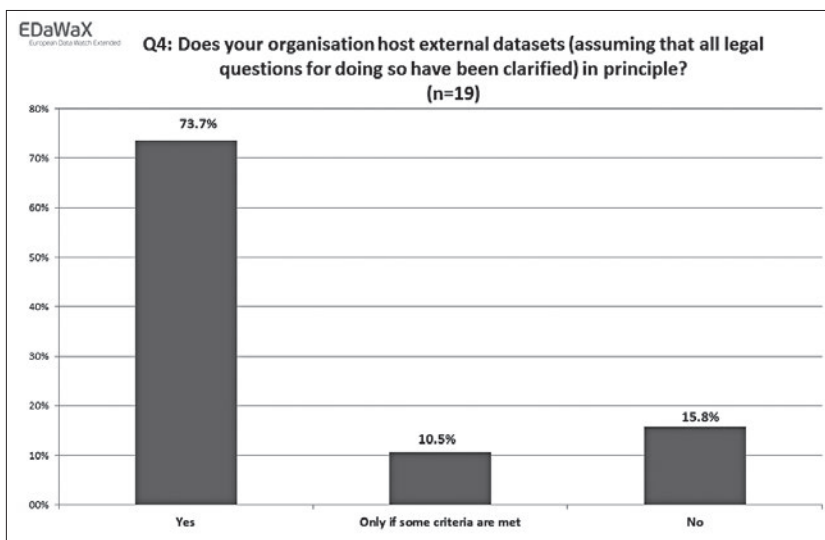


Abb. 3: Bereitschaft zum Hosting von extern erzeugten Forschungsdatensätzen durch die jeweiligen Einrichtungen

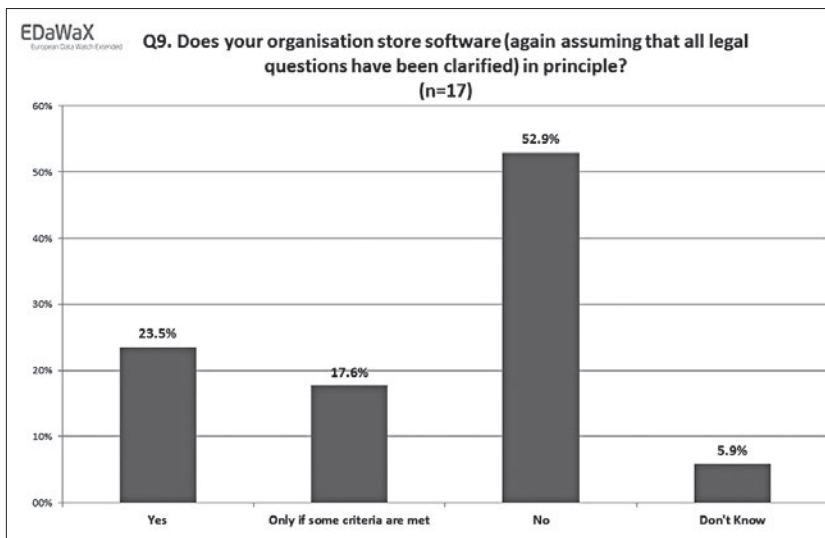


Abb. 4: Bereitschaft zur Speicherung von (selbstgeschriebener) Software

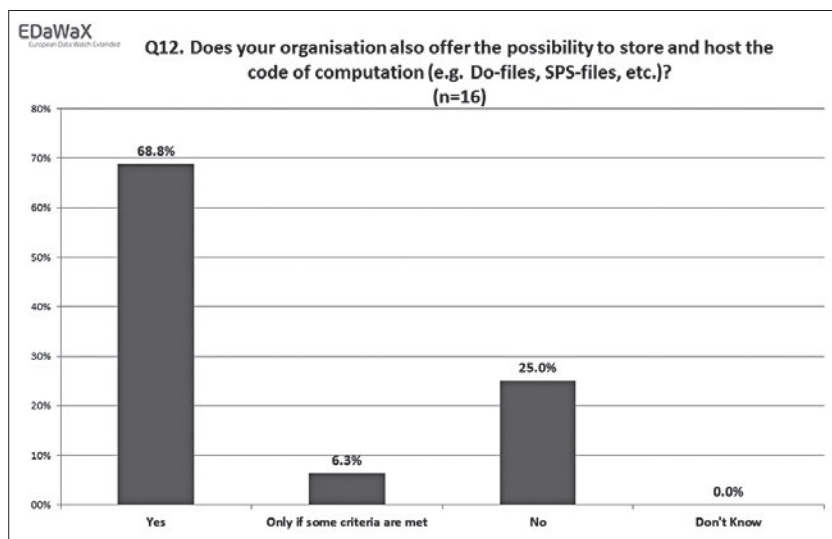


Abb. 5: Bereitschaft zur Speicherung und zum Hosting des Berechnungscodes (Syntax)

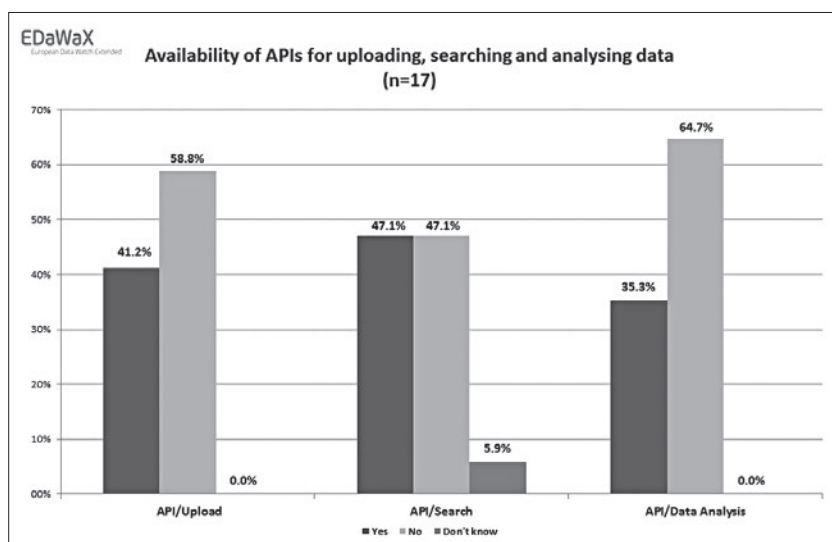


Abb. 6: Verfügbarkeit von Programmierschnittstellen (APIs) nach Auskunft der Befragten

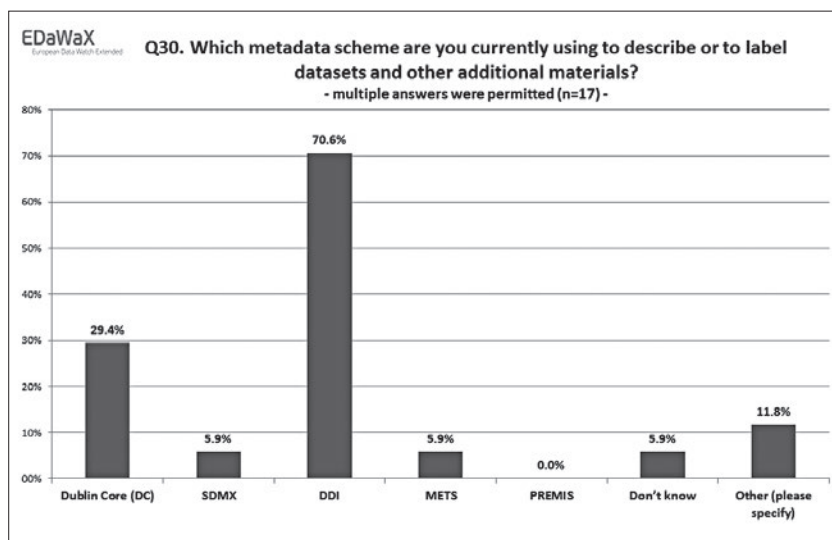


Abb. 7: Genutzte Metadatenschemata zur Beschreibung von Datensätzen und weiteren Materialien

APIs

Im Zuge der Untersuchung wurde auch die Verfügbarkeit von Programmierschnittstellen (sogenannten APIs) thematisiert, mit denen ein automatisierter Austausch von Daten ermöglicht wird und die für einen dezentralen Einsatz von Softwareapplikationen für das Management solcher Forschungsdaten relevant sind.

Die Befragung ergab, dass weniger als die Hälfte aller Organisationen angab, über solche Schnittstellen zu verfügen. Am häufigsten wurden APIs zur Suche von Datensätzen genannt (47 %), gefolgt von APIs für den Upload von Forschungsdaten. Etwas mehr als ein Drittel (35 %) der Antwortenden gibt zudem an, über Schnittstellen für die Analyse von Forschungsdaten zu verfügen. Eine im Rahmen des EDaWaX-Projekts durchgeführte Analyse der angegebenen Schnittstellen ergab jedoch, dass es sich bei diesen Schnittstellen durchgängig nur um Suchmasken auf den jeweiligen Webseiten handelt. Schnittstellen im Sinne eines extern möglichen Lese- und Schreibzugriffs sind daher als weitgehend unbekannt zu charakterisieren.

Metadaten und Metadatenerstellung

Erfragt wurden auch die Metadatenschemata, die die befragten Organisationen bei ihrer Arbeit verwenden. Dabei zeigte sich, dass mehr als 70 % der Antwortenden DDI verwenden. Wesentlich seltener wurde DC (knapp 30 %) genannt.¹⁴ Alle anderen Metadatenschemata werden nur vereinzelt genutzt.

Persistente Identifier (PI)

Von Interesse war zudem, ob persistente Identifikatoren (wie handle, DOI, URN, etc.) in den befragten Organisationen Verwendung finden. Die persistente Identifikation von Forschungsdaten ist u.a. wichtig für die Zitierbarkeit von Forschungsdatensätzen. Organisationen aus dem Sample vergaben solche Identifier standardmäßig in mehr als 56 % der Fälle, fast ein Drittel vergab solche Identifier jedoch nicht.

Unterstützung von Semantic Web Technologien

In der Befragung wurde auch nach der Verwendung von RDF gefragt – einem Datenmodell zur Beschreibung von Ressourcen mittels semantischer Technologien. Von den antwortenden Organisationen gab nur eine Minderheit von etwa 6 % an, RDF-Daten bereitzustellen. Fast ein Viertel der Antwortenden machte hierzu keine Angaben, was teilweise auf einen geringen Bekanntheitsgrad von RDF zurückzuführen sein dürfte.

Support bei der Metadatenerstellung

Die Achillesferse für die Nachnutzbarkeit von Forschungsdaten ist häufig die Qualität der Dokumentation von Forschungsdaten. Daher war es von besonderem Interesse zu erfahren, ob und wie die untersuchten Organisationen Forscher/innen bei der Generierung von Metadaten unterstützen.

Die Befragung ergab, dass die Mehrheit der antwortenden Organisationen (fast 65 %) über einen entsprechenden Support für Wissenschaftler/innen verfügt. Dabei war auch von Interesse, ob es softwarebasierte Unterstützung für die Erstellung von Metadaten an den Institutionen gibt, wie dies beispielsweise durch entsprechende Eingabemasken und die Konvertierung der eingegebenen Inhalte in standardisierte Metadaten der Fall sein kann. Hier zeigte sich, dass über 35 % der Antwortenden über einen solchen softwarebasierten Support verfügen. Auffällig ist die Zahl an Nennungen im Bereich *other*. Hier wurden beispielsweise auch schriftlich auszufüllende *Data Deposit Forms* aufgeführt. Auf Nachfrage nach dem Namen bzw. der Art der Software, gaben mindestens zwei Institutionen an, dass sie Nesstar¹⁵ einsetzen. Viele Organisationen nutzen zudem Eigenentwicklungen.

Langzeitarchivierung

In der Befragung war ebenfalls von Interesse, in welchem Umfang Maßnahmen zur Langzeitarchivierung von Forschungsdaten getroffen werden. Es wurde nach »spezifischen Maßnahmen« zur Langzeitarchivierung gefragt und als Beispiel die Formatmigrationen als eine der derzeit favorisierten Maßnahmen zum Langzeiterhalt digitaler Bestände erfasst. Mehr als 80 % der Antwortenden gaben die Rückmeldung, dass ihre Institutionen entsprechende Maßnahmen zur Langzeitarchivierung umsetzen.

FAZIT

Die durchgeführte Desktop-Analyse und insbesondere die Befragungsergebnisse zeigen, dass sozial- und wirtschaftswissenschaftliche Forschungsdatenzentren relevante Speicherorte für publikationsbezogene Forschungsdaten sein können, da sie verschiedene Voraussetzungen dafür bereits erfüllen. Dennoch gibt es unter den 22 Organisationen, die Angaben machten, bislang keine Institution, die in Gänze alle Anforderungen hinsichtlich Speicherung und Hosting solcher publikationsbezogenen Forschungsdaten in den Sozial- und Wirtschaftswissenschaften erfüllt.

Zusammenfassend lassen sich folgende Ergebnisse anführen:

- Etwa drei Viertel aller antwortenden Einrichtungen akzeptieren grundsätzlich externe Forschungsda-

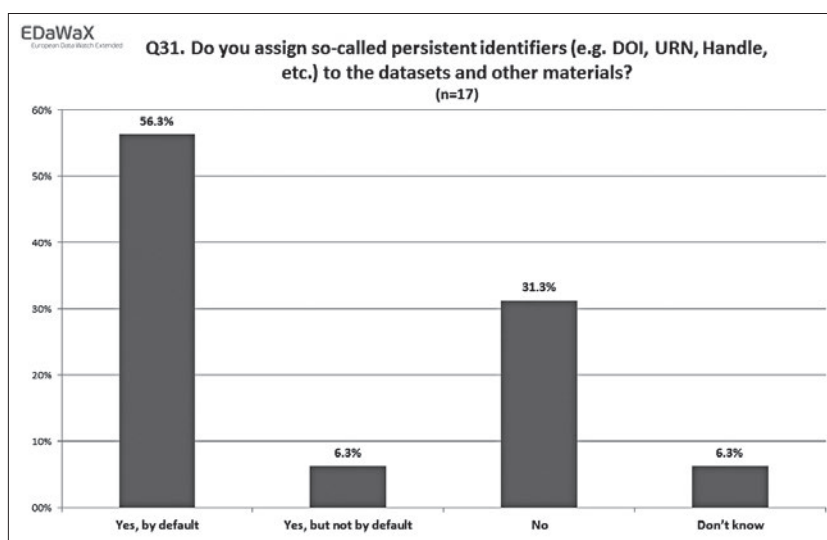


Abb. 8: Vergabe von Persistenten Identifiern bei den befragten Einrichtungen

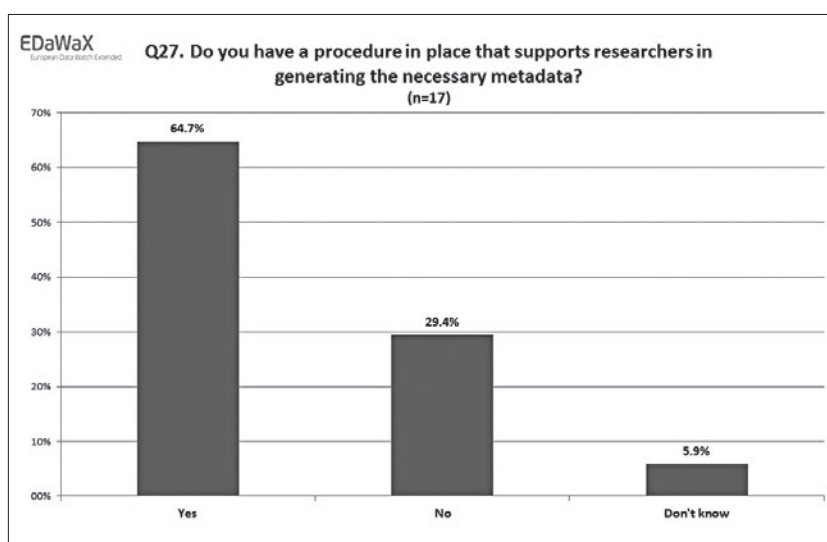


Abb. 9: Allgemeine Unterstützung von Forschenden bei der Metadategenerierung

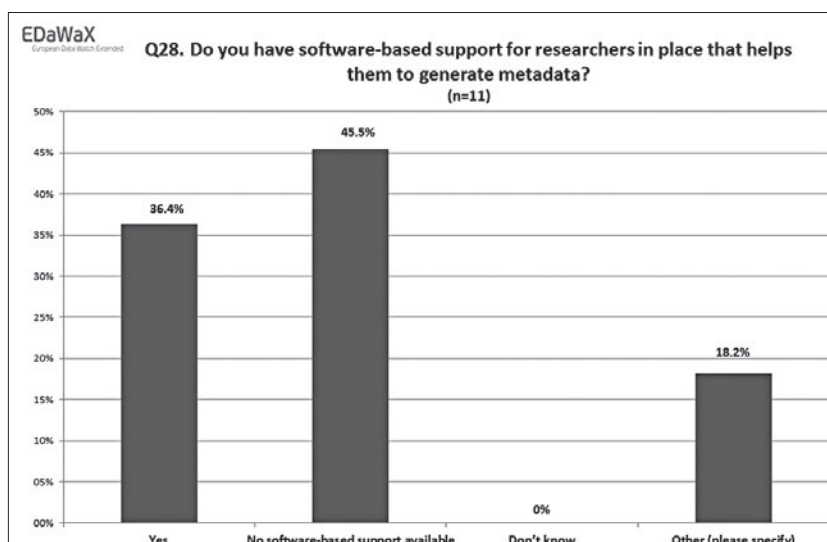


Abb. 10: Einrichtungen mit softwarebasierten Verfahren zur Unterstützung der Metadatenerstellung

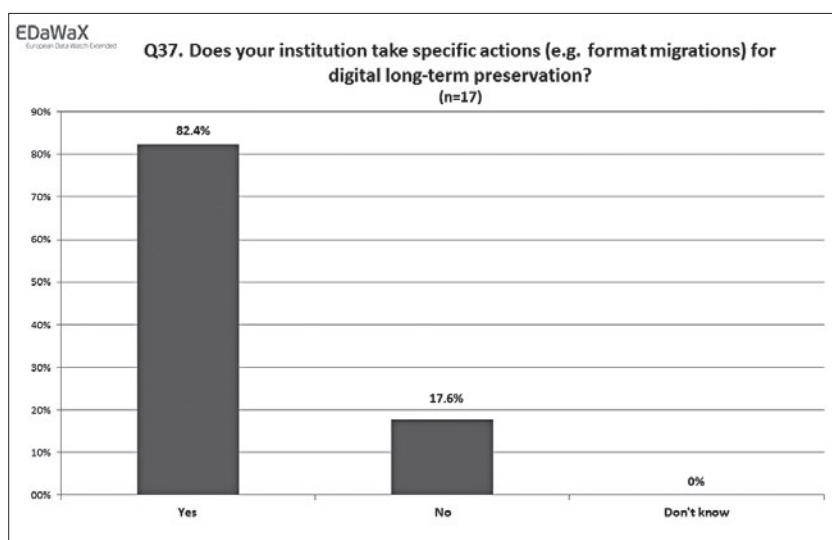


Abb. 11: Umsetzung von Maßnahmen zur Langzeitarchivierung (z. B. Format-Migrationen) nach Auskunft der Befragten

ten, inklusive publikationsbezogener Forschungsdaten. Allerdings gibt es z.T. Einschränkungen, etwa aufgrund der fachlichen oder regionalen Zuständigkeit oder hinsichtlich der qualitativen Anforderungen an solche Datensätze.

- Fast ebenso hoch (annähernd 75 %) ist die Anzahl der Datenzentren, die den zugehörigen Berechnungscode (Syntax) der abgegebenen Berechnungen prinzipiell speichern und hosten. Falls zur Berechnung empirischer Ergebnisse spezielle (selbstgeschriebene) Software verwendet wurde, wird diese allerdings nur von etwa 40 % der antwortenden Organisationen für Speicherung und Hosting akzeptiert.
- An eingesetzten Metadatenschemata dominiert klar DDI (70 %) vor Dublin Core (30 % – Mehrfachnennungen möglich). Knapp zwei Drittel zeichnet die Datensätze zudem mit Persistenten Identifikatoren aus und macht sie so leichter zitierfähig. Etwa drei Viertel aller Befragten leistet zudem Unterstützung bei der Eingabe der Metadaten durch Forschende.
- Schnittstellen für die externe Suche oder den Upload von Datensätzen werden bislang nicht durch die befragten Einrichtungen angeboten. Kaum verbreitet ist auch der Einsatz von semantischen Technologien wie z.B. RDF.
- Die antwortenden Organisationen sorgen zudem zu über 80 % für die Langzeitverfügbarkeit der bei ihnen gehosteten Datensätze.

FDZs können Speicherung und Hosting übernehmen

FDZs und Bibliotheken sollen eng kooperieren

Aus den erzielten Ergebnissen der Befragung kann schlussgefolgert werden, dass Forschungsdatenzen-

tren und Bibliotheken beim Management von publikationsbezogenen Forschungsdaten eng zusammenarbeiten sollten. Beide Typen von Einrichtungen haben spezifische Erfahrungen und Kompetenzen, die sich gut bei der Entwicklung und Etablierung publikationsbezogener Datenarchive ergänzen können. Diese speziellen Archive sollten für jede einzelne Publikation die ihr zugrundeliegenden Daten in nutzerfreundlicher Form für Re-Analysen zugänglich machen.

Für die Entwicklung und den technischen Ausbau der im EDaWaX-Projekt entwickelten Pilotapplikation für das Forschungsdatenmanagement in (wirtschaftswissenschaftlichen) Fachzeitschriften strebt das Projekt einen Workflow an, der ebenfalls auf den komplementären Kompetenzen von Bibliotheken und Forschungsdatenzentren beruht: So übernimmt die ZBW – Leibniz-Informationszentrum Wirtschaft – in dieser Arbeitsteilung die Aufgabe des Hostings und der dauerhaften Pflege des Metadatenkataloges zu den gespeicherten Forschungsdaten. Auch die Implementierung von Schnittstellen (APIs) zu anderen Bibliotheks- oder Forschungsdatenkatalogen ist eine Aufgabe, für die die Bibliothek sich verantwortlich zeigt, da sie hier bereits auf zahlreiche Erfahrungen zurückblicken kann. Eines der vielgenutzten deutschen Forschungsdatenzentren, das FDZ SOEP (Sozio-oekonomisches Panel), wird als eine beispielhafte Pilot-Institution in naher Zukunft komplementär die Aufgaben des Hostings, der Speicherung und der Langzeitarchivierung der Forschungsdaten übernehmen, die durch die Anwendung an die Fachzeitschriften übermittelt wurden. Dieses Zusammenspiel von Forschungsdatenzentren und Bibliotheken verdeutlicht anschaulich die Synergieeffekte, die durch das Zusammenwirken von wissenschaftlichen Informationsdienstleistern mit komplementären Kompetenzen erzielt werden können.

LITERATUR

- Anderson, R.; Greene, W. H.; McCullough, B. D.; Vinod, H. D. (2008):** The Role of Data/Code Archives in the Future of Economic Research. In: Journal of Economic Methodology, 15 (1), S. 99–119.
- Andreoli-Versbach P.; Mueller-Langer, F. (2013):** Open Access to Data : An Ideal Professed but not Practised, RatSWD Working Paper Series, Nr. 215, Berlin. Verfügbar unter: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2224146
- De Cock Bruning, M.; van Dither, B.; Jeppersen de Boer, C. G.; Ringnald, A. (2011):** The legal status of research data in the Knowledge Exchange partner countries. Verfügbar unter: www.knowledge-exchange.info/Default.aspx?ID=461

Häder, M. (2009): Der Datenschutz in den Sozialwissenschaften. Anmerkungen zur Praxis sozialwissenschaftlicher Erhebungen und Datenverarbeitung in Deutschland. RatSWD Working Paper Series (90). Verfügbar unter: www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_90.pdf

Hillegeist, T. (2012): Rechtliche Probleme der elektronischen Langzeitarchivierung wissenschaftlicher Primärdaten, Göttinger Schriften zur Internetforschung (8). Verfügbar unter: http://webdoc.sub.gwdg.de/uni-verlag/2012/GSI8_Hillegeist.pdf

Huschka, D.; Oellers, C.; Ott, N.; Wagner, G. G. (2012): Datenmanagement und Data Sharing : Erfahrungen in den Sozial- und Wirtschaftswissenschaften. In: R. Altenhöner; C. Oellers (Hrsg.): Langzeitarchivierung von Forschungsdaten – Standards und disziplinspezifische Lösungen, Berlin: Scivero, S. 23–36. Verfügbar unter: www.ratswd.de/dl/downloads/langzeitarchivierung_von_forschungsdaten.pdf

King, G. (1995): Replication, replication. In: PS: Political Science and Politics, 28, S. 443–499. Verfügbar unter: <http://gking.harvard.edu/gking/files/replication.pdf>

McCullough, B. D. (2009): Open Access Economics Journals and the Market for Reproducible Economic Research. In: Economic Analysis and Policy, 39 (1), S. 117–126.

Research Information Network (2011): Data centres : their use, value and impact. A Research Information Network report. September 2011. Verfügbar unter: www.rin.ac.uk/system/files/attachments/Data_Centres_Report.pdf

Spindler, G.; Hillegeist T. (2008): KoLaWiss-Gutachten AP 4 : Recht, Rechtsexpertise für das Projekt »Kooperative Langzeitarchivierung an Wissenschaftsstandorten« (KoLaWiss). Verfügbar unter: http://kolawiss.uni-goettingen.de/projektergebnisse/AP4_Report.pdf

Siebert, O.; Toepfer R.; Vlaeminck, S. (2012): Forschungsdatenmanagement in den Wirtschaftswissenschaften – Ausgewählte Dienste und Projekte der Deutschen Zentralbibliothek für Wirtschaftswissenschaften – Leibniz-Informationszentrum Wirtschaft (ZBW). In: R. Altenhöner; C. Oellers (Hrsg.): Langzeitarchivierung von Forschungsdaten – Standards und disziplinspezifische Lösungen, Berlin, Scivero, S. 217–225. Verfügbar unter: www.ratswd.de/dl/downloads/langzeitarchivierung_von_forschungsdaten.pdf

Vlaeminck, S. (2013): Data Management in Scholarly Journals and possible Roles for Libraries – Some Insights from EDaWaX. In: LIBER Quarterly, 23 (1). Verfügbar unter: <http://liber.library.uu.nl/index.php/lq/article/view/URN%3ANBN%3ANL%3AUI%3A10-1-114595>

Vlaeminck, S.; Siebert, O. (2012): Welche Rolle spielen Forschungsdaten eigentlich für Fachzeitschriften?

PUBLIZISTIK IN JÜDISCHEN
DISPLACED-PERSONS-CAMPS IM
NACHKRIEGSDEUTSCHLAND
Charakteristika, Medientypen und
bibliothekarische Überlieferung
Hrsg. von Anne-Katrin Henkel
und Thomas Rahe
2014. 194 Seiten, gebunden, Fadenheftung
ISBN 978-3-465-04216-7
ZfBB Sonderband 112

Die Geschichte der jüdischen Displaced Persons stand lange im Schatten der Shoah-Forschung. Dabei waren es die Überlebenden der Shoa, die nun großenteils in den DP-Camps vor allem auf deutschem Boden lebten, die die ersten Impulse gaben für die Dokumentation der von ihnen selbst erlebten nationalsozialistischen Verfolgung. Die frühen Bemühungen einer Dokumentation der Shoa »von unten« sind ein zentrales Charakteristikum des Lebens in den jüdischen DP-Camps im Nachkriegsdeutschland. Diese Erkenntnis war der Anlass für die Gottfried Wilhelm Leibniz-Bibliothek, die im DP-Camp Bergen-Belsen zwischen 1945 und 1950 (re-)produzierte oder vom Zentralkomitee der befreiten Juden auf verschlungenen Wegen erworbene jüdische Buchkultur sukzessive durch eine gezielte Erwerbung zu rekonstruieren.

Vor diesem Hintergrund entstand dieser Band. Das Spektrum der Beiträge reicht von der bibliothekarischen Überlieferung und ihren Bedingungsfaktoren, ihren formalen und inhaltlichen Charakteristika, den Akteuren und Kontexten der Publikationstätigkeit, den Funktionen dieser Publikationen in den jüdischen DP-Camps, ihrer historischen und historiographischen Bedeutung bis hin zu ihrer Stellung innerhalb der Kontinuität deutsch-jüdischer Geschichte.



VITTORIO KLOSTERMANN

ten? Eine Analyse mit Fokus auf die Wirtschaftswissenschaften. In: RatSWD Working Papers, Nr. 210, Berlin. Verfügbar unter: www.ratswd.de/download/RatSWD_WP_2012/RatSWD_WP_210.pdf

Vlaeminck, S.; Wagner, G. G.; Wagner, J.; Harhoff, D.; Siegert, O. (2013): Replizierbare Forschung in den Wirtschaftswissenschaften erhöhen. In: LIBREAS. Library Ideas, 23: Forschungsdaten. Metadaten. Noch mehr Daten. Forschungsdatenmanagement. Verfügbar unter: <http://edoc.hu-berlin.de/libreas/23/vlaeminck-sven-1/PDF/vlaeminck.pdf> (urn:nbn:de:kobv:11-100212694)

¹ Die ALLBUS (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften) wird jährlich von der GESIS – Leibniz Institut für Sozialwissenschaften (Köln und Mannheim) durchgeführt. Siehe www.gesis.org/de/allbus

² Das Sozio-oekonomische Panel (SOEP) ist eine repräsentative Wiederholungsbefragung, die seit mehr als 25 Jahren durchgeführt wird. Im Auftrag des DIW Berlin werden jedes Jahr in Deutschland über 20.000 Personen aus rund 11.000 Haushalten befragt. Siehe www.diw.de/de/diw_02.c.221178.de/ueber_uns.html

³ Die hier vorgestellten Projektergebnisse wurden im Rahmen des Projekts EDaWaX (European Data Watch Extended, www.edawax.de) erarbeitet. EDaWaX wird von der Deutschen Forschungsgemeinschaft (DFG) gefördert. Institutionell beteiligt sind an diesem Projekt der Rat für Sozial- und Wirtschaftsdaten (RatSWD), das Institut Innotec der LMU München in Kooperation mit dem Max Planck Institute for Intellectual Property and Competition Law (IMPRS-CI) sowie die Deutsche Zentralbibliothek für Wirtschaftswissenschaften (ZBW). Neben den Autoren sind folgende Personen am EDaWaX Projekt beteiligt: Prof. Klaus Toepfer (ZBW), Prof. Joachim Wagner, Prof. Dietmar Harhoff (Innotec und MCIER), Dr. Brigitte Preissl (ZBW), Patrick Andreoli-Versbach (IMPRS-CI), Dr. Frank Mueller-Langer (IMPRS-CI und MCIER), Ralf Toepfer (ZBW) und Dr. Hendrik Bunke (ZBW).

⁴ Zwar existieren verschiedene Publikationen und einzelne Rechtsexperten zum Umgang mit Forschungsdaten, die jedoch etwaige Unsicherheiten auf Seiten der Forschenden in Bezug auf ein mögliches Data Sharing nicht hinreichend klar beseitigen dürften. Vgl. u. a. Häder, M. (2009): Der Datenschutz in den Sozialwissenschaften. Anmerkungen zur Praxis sozialwissenschaftlicher Erhebungen und Datenverarbeitung in Deutschland. RatSWD Working Paper Series (90). Verfügbar unter: www.ratswd.de/download/RatSWD_WP_2009/RatSWD_WP_90.pdf; Hillegeist, T. (2012): Rechtliche Probleme der elektronischen Langzeitarchivierung wissenschaftlicher Primärdaten, Göttinger Schriften zur Internetforschung (8). Verfügbar unter: http://webdoc.sub.gwdg.de/univverlag/2012/GSI8_Hillegeist.pdf (insbesondere Kapitel A); Spindler, G.; Hillegeist T. (2008): KoLaWiss-Gutachten AP 4: Recht, Rechtsexpertise für das Projekt »Kooperative Langzeitarchivierung an Wissenschaftsstandorten« (KoLaWiss). Verfügbar unter: http://kolawiss.uni-goettingen.de/projektergebnisse/AP4_Report.pdf sowie De Cock Bruning, M.; van Dither, B.; Jeppersen de Boer, C. G.;

Ringnalda, A. (2011): The legal status of research data in the Knowledge Exchange partner countries. Verfügbar unter: www.knowledge-exchange.info/Default.aspx?ID=461

⁵ Webseite von Dataverse, www.thedata.org/

⁶ Rat für Sozial- und Wirtschaftsdaten, www.ratswd.de

⁷ Council of European Social Science Data Archives, www.cessda.org

⁸ Inzwischen wurde das »Publication Related Archive« des ICPSR in »Replication Datasets« umbenannt.

⁹ Eine Übersicht der Journals, in denen Artikel erschienen sind, die auf Basis der hinterlegten Daten entstanden sind, findet sich unter www.icpsr.umich.edu/icpsrweb/ICPSR/biblio/journals?collection=DATA

¹⁰ Die Webseite von DANS EASY ist erreichbar unter <https://easy.dans.knaw.nl/ui/home>

¹¹ Hilfreiche Informationen finden sich beispielsweise in dem Dokument »Deposit instructions for social and behavioural sciences« unter www.dans.knaw.nl/sites/default/files/file/EASY/Deponeerinstructie%20MaGw%20UK%20DEF.pdf

¹² Die Befragten hatten zudem die Möglichkeit, direkt aus dem Fragebogen heraus Beispiele für solche publikationsbezogenen Forschungsdaten und die einzelnen Bestandteile einer kompletten Einreichung anzusehen.

¹³ Vgl. hierzu beispielhaft die Policy der American Economic Review (AER) unter www.aeaweb.org/aer/data.php

¹⁴ In der Befragung selbst wurden die Forschungsdatenzentren auch nach der Verwendung von XML befragt. Da XML jedoch kein Metadatenschema ist, sondern ein Verfahren zur Repräsentation von Metadaten, wurden die Ergebnisse hier nicht weiter berücksichtigt.

¹⁵ Webseite von nesstar: www.nesstar.com

DIE VERFASSER

Sven Vlaeminck, Projektmanager European Data Watch Extended (EDaWaX), Deutsche Zentralbibliothek für Wirtschaftswissenschaften – Leibniz-Informationszentrum Wirtschaft (ZBW), Neuer Jungfernstieg 21, 20354 Hamburg, Tel.: 040 – 42834-415, E-Mail: s.vlaeminck@zbw.eu

Prof. Dr. Gert G. Wagner, Rat für Sozial- und Wirtschaftsdaten (RatSWD), DIW Berlin (Dt. Institut für Wirtschaftsforschung), TU Berlin und MPI für Bildungsforschung, Mohrenstraße 58, 10117 Berlin, Tel.: 030 – 89789-290, E-Mail: gwagner@diw.de