

225

Zur Rolle von Forschungsdaten zentren beim Management von publikationsbezogenen Forschungsdaten

Ergebnisse einer Befragung
von wissenschaftlichen
Infrastrukturdienstleistern

Sven Vlaeminck und Gert G. Wagner

Oktober 2013



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 erschienenen 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)

Zur Rolle von Forschungsdatenzentren beim Management von publikations- bezogenen Forschungsdaten

Ergebnisse einer Befragung von wissenschaftlichen
Infrastrukturdienstleistern

Sven Vlaeminck

Deutsche Zentralbibliothek für Wirtschaftswissenschaften – Leibniz Informationszentrum Wirtschaft (ZBW), Hamburg

Gert G. Wagner

Rat für Sozial- und Wirtschaftsdaten (RatSWD), TU Berlin, MPI für Bildungsforschung und DIW Berlin

Die hier vorgestellten Projektergebnisse wurden im Rahmen des Projekts EDaWaX (European Data Watch Extended, www.edawax.de) erarbeitet. EDaWaX wird von der Deutschen Forschungsgemeinschaft (www.dfg.de) gefördert. Institutionell beteiligt sind an diesem Projekt der Rat für Sozial- und Wirtschaftsdaten (RatSWD), das Institut Inno-tec 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 Tochtermann (ZBW), Prof. Joachim Wagner, Prof. Dietmar Harhoff (Inno-tec 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).

Zusammenfassung

Im vorliegenden Papier werden die Ergebnisse einer Analyse zusammengefasst, in der untersucht wurde, welche Services für das Management von publikationsbezogenen Forschungsdaten gegenwärtig bei wissenschaftlichen Infrastrukturdienstleistern bestehen. Mit Hilfe einer Befragung und einer Desktop-Recherche wurde ermittelt, dass etwa Dreiviertel aller untersuchten einschlägigen Forschungsdatenzentren, Archive und Bibliotheken grundsätzlich externe Forschungsdaten, inklusive publikationsbezogener Forschungsdaten, archivieren. 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 Befragten für Speicherung und Hosting akzeptiert. Die befragten Organisationen sorgen zudem zu über 80% für die Langzeitverfügbarkeit der bei ihnen gehosteten Datensätze. 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 Dreiviertel 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.

Summary

This paper summarizes the findings of an analysis among scientific infrastructure service providers. These service providers have been evaluated in regard to their potential services for the management of publication-related research data. By conducting a desk research and an online survey, we found out that almost three quarters of all responding research data centres, archives and libraries generally store externally generated research data – what also applies to publication-related data. Almost 75% of all respondents also store and host the code of computation (the syntax of statistical analyses). If self-written software components have been used to generate research outputs, only 40% of all respondents accept these software components for storing and hosting. Eight in ten institutions also stated that they are taking specific actions for digital long-term preservation of their data. In regard to the documentation of stored and hosted research data almost 70% of all respondents claimed to use the metadata schema of the Data Documentation Initiative (DDI); Dublin Core was used by 30 percent (multiple answers were permitted). Almost two thirds also used persistent identifiers to facilitate citation of these datasets. Three in four respondents also stated to support researchers in creating metadata for their data. Application programming interfaces (APIs) for uploading or searching datasets currently have not been implemented by any of the respondents yet. Little widespread is the use of semantic technologies like RDF.

An English version of this paper is available as RatSWD Working Paper No. 226.

JEL Classification: C81, C88, H42, H54

Keywords: Research Data Centers, Libraries, Archives, Research Data Management, Journals, Replicability

Hintergrund

In den Wirtschaftswissenschaften finden sich zunehmend mehr Publikationen in Fachzeitschriften, in denen Forschende selbst erstellte oder extern verfügbare Datensätze nach eigenen Fragestellungen ausgewertet haben.

Die für diese Forschungsfragestellungen genutzten Daten stammen aus unterschiedlichsten Quellen. Im Unterschied zu stärker empirisch ausgerichteten Wissenschaftsdisziplinen werden jedoch in der Regel in den Wirtschaftswissenschaften seltener eigene Forschungsdatensätze erstellt. Ausnahme ist hier die experimentelle Wirtschaftsforschung, die der häufig Datensätze als Ergebnis spieltheoretisch motivierter Forschung erzeugt werden. Jedoch werden diese Datensätze in der Regel nicht für Dritte dokumentiert und archiviert. Stattdessen greifen Wirtschaftsforscher/innen oftmals auf Daten der amtlichen Statistik zurück oder auf Studien, die durch spezialisierte Forschungseinheiten erhoben werden (z. B. der ALLBUS der GESIS¹ oder das SOEP² am DIW Berlin). Häufig werden auch relevante Datensätze von Firmen, wie etwa Bloomberg oder Thomson Reuters gekauft.

Doch obwohl zunehmend mehr Publikationen auf Basis der Auswertung von Daten erscheinen, gibt es in den Wirtschaftswissenschaften (wie in den meisten anderen wissenschaftlichen Disziplinen) bislang jedoch kaum wirklich effektive Möglichkeiten, die in Fachzeitschriften publizierten Forschungsergebnisse im Kontext der zugehörigen Artikel zu replizieren, zu prüfen oder für eine Nachnutzung und zur Unterstützung des wissenschaftlichen Diskurses bereit zu stellen.

Selbst solche Forschungsdaten, die auf allgemein zugänglichen Datensätzen (z. B. ALLBUS oder SOEP) beruhen, werden in der Regel nicht in ihrer spezifischen Auswahl und Bereinigung (als finales Arbeits-File) archiviert. Damit sind Replikationen zwar nicht ausgeschlossen, aber eine Replikation fällt bei anspruchsvollen Analysen, die auf spezifischer Auswahl und Bereinigungen beruhen, schwer.

Die derzeitige Situation stellt sowohl die Fachdisziplin als auch wissenschaftliche Infrastrukturdienstleister wie Bibliotheken und Datenzentren vor Herausforderungen.

¹ Der ALLBUS (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften) wird jährlich von der GESIS - Leibniz Institut für Sozialwissenschaften (Köln und Mannheim) durchgeführt. Siehe <http://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 http://www.diw.de/de/diw_02.c.221178.de/ueber_uns.html

Weshalb ist wirtschaftswissenschaftliche Forschung oft nicht replizierbar?

Die Gründe für die fehlende Replizierbarkeit wirtschaftswissenschaftlicher Forschung lassen sich auf verschiedenen Ebenen verorten:

- 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 Data Sharing, die oftmals sehr zeitintensiv ist, nicht. Dies steht im krassen 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, 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 Schiefelage 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).³

³ Zwar existieren verschiedene Publikationen und einzelne Rechtsexpertisen 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: http://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/univerlag/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: <http://www.knowledge-exchange.info/Default.aspx?ID=461>

- Derzeit verfügen nur wenige wirtschaftswissenschaftliche Fachzeitschriften über Richtlinien, die den Umgang mit den verwendeten Forschungsdaten im Detail regeln (McCullough, 2009): So genannte Data Availability Policies beispielsweise 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 bereit zu stellen. Solche Richtlinien folgen somit oftmals dem vom Gary King 1995 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 Dataverse⁴, 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 was für) Services angeboten werden.

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 et al., 2013). Verschiedene Publikationen aus dem Projektkontext befassen sich zudem mit Teilaspekten des Projekts: So beleuchtet 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).

Dieses Working Paper beschreibt ergänzend die Ergebnisse einer Evaluierung von wissenschaftlichen Infrastrukturdienstleistern in Bezug auf deren Services für das Management von publikationsbezogenen Forschungsdaten in den Sozial- und Wirtschaftswissenschaften.

⁴ Webseite von Dataverse, <http://www.thedata.org/>

Bieten Forschungsdatenzentren eigentlich 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 sozial- 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).

Im Rahmen des EDaWaX-Projekts wurde daher eine Studie durchgeführt, die evaluierte, ob Services für solche publikationsbezogenen Forschungsdaten bei Infrastrukturdienstleistern wie Forschungsdatenzentren (FDZ), Bibliotheken oder auch Archiven existieren.

Das Projekt erstellte dazu zunächst eine Liste von 46 Organisationen aus dem Bereich der wissenschaftlichen Infrastrukturdienstleister. Darunter fanden sich sowohl die vom Rat für Sozial- und Wirtschaftsdaten (RatSWD)⁵ akkreditierten Forschungsdaten- und Datenservicezentren, die im CESSDA-Verbund⁶ organisierten zumeist europäischen Datenzentren, Bibliotheksverbünde sowie einzelne Bibliotheken und Archive.

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)⁷ vorliegt, was bereits von zahlreichen Autor/inn/en für die Hinterlegung von publikationsbezogenen Forschungsdaten genutzt wurde.⁸ Das niederländische DANS EASY⁹ bietet zwar keine dezidierten Services für publikationsbezogenen Forschungsdaten an, jedoch ist es prinzipiell auch möglich DANS EASY für die Hinterlegung solcher Daten zu nutzen.¹⁰

⁵ Rat für Sozial- und Wirtschaftsdaten, www.ratswd.de

⁶ Council of European Social Science Data Archives, www.CESSDA.org

⁷ Inzwischen wurde das „Publication Related Archive“ des ICPSR in „Replication Datasets“ umbenannt.

⁸ Eine Übersicht der Journals in denen Artikel erschienen sind, die auf Basis der hinterlegten Daten entstanden sind, findet sich unter

<http://www.icpsr.umich.edu/icpsrweb/ICPSR/biblio/journals?collection=DATA>

⁹ Die Webseite von DANS EASY ist erreichbar unter: <https://easy.dans.knaw.nl/ui/home>

¹⁰ Hilfreiche Informationen finden sich beispielsweise in den Dokument „Deposit instructions for social and behavioural sciences“ unter

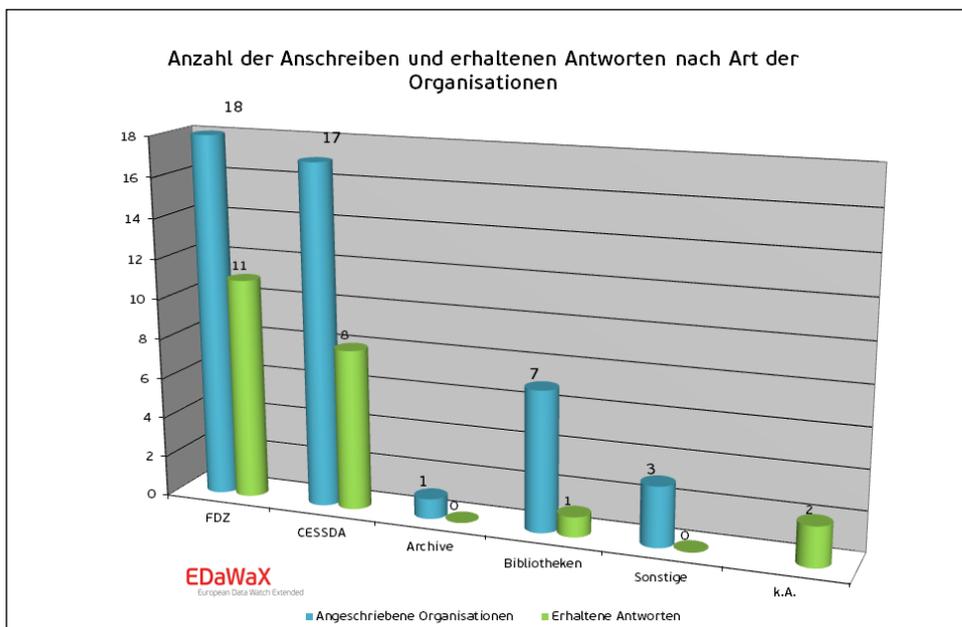
<http://www.dans.knaw.nl/sites/default/files/file/EASY/Deponerinstruactie%20MaGw%20UK%20DEF.pdf>.

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 36 nationale und internationale Forschungsdatenzentren (FDZs) und Datenservicezentren (DSZs), 1 Archiv, 7 Bibliotheksverbände 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 Bibliotheksverbände und Archive, aber auch die drei unter „Sonstiges“ berücksichtigten außer-europäischen Forschungsdatenzentren.



Für die Bibliotheksverbände 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.

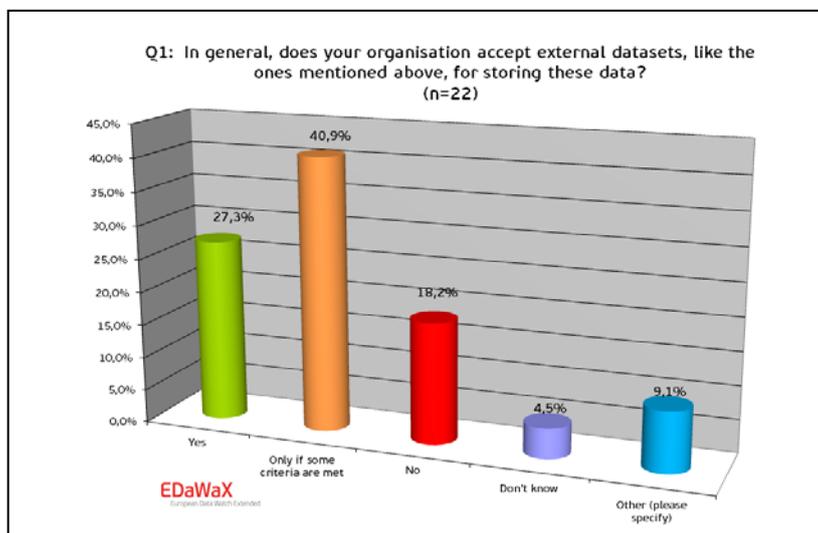
Inhaltliche Auswertung

Zunächst wurde mit der Befragung untersucht, ob die angeschriebenen Institutionen publikationsbezogene Forschungsdaten prinzipiell speichern und hosten.¹¹ 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.¹²

Datensätze

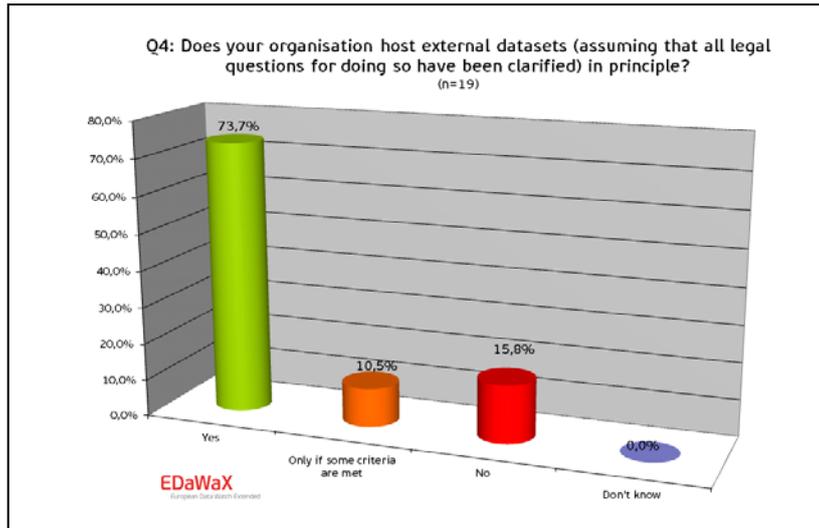
Von den untersuchten Organisationen akzeptierten mehr als drei Viertel externe Datensätze zur Speicherung. Der größte Anteil der Befragten gab dabei an, solche Forschungsdaten nur anzunehmen, wenn gewisse Kriterien erfüllt sind. Solche Kriterien bestehen etwa in Form der spezifischen thematischen Zuständigkeiten der deutschen FDZs, aber auch in Form von regionalen, überregionalen und/oder fachspezifische Zuständigkeiten. Zudem wurden technisch-organisatorische Aspekte (Dokumentation, Maschinenlesbarkeit) und rechtliche Fragestellungen als Kriterien genannt.

Von den befragten 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.



¹¹ Die Befragten hatten zudem die Möglichkeit, direkt aus dem Fragebogen heraus Beispiele für solche publikationsbezogenen Forschungsdaten und die einzelnen Bestandteile einer kompletten Einreichung anzusehen.

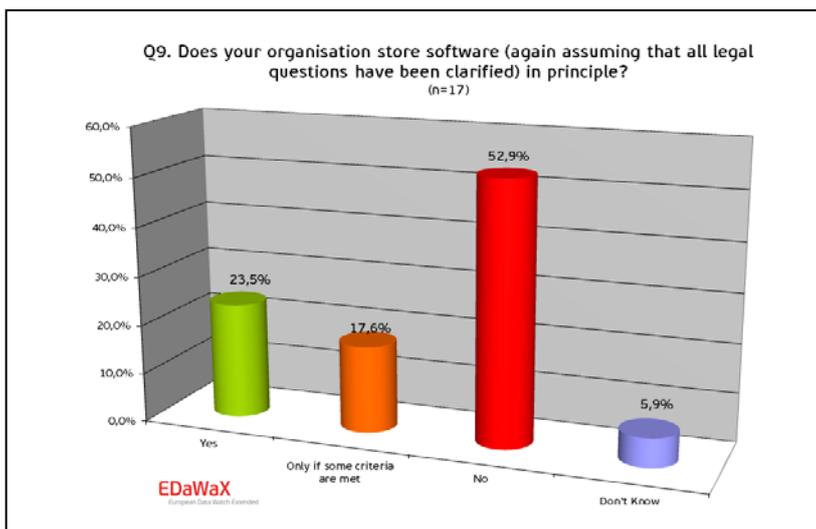
¹² Vgl. hierzu beispielhaft die Policy der American Economic Review (AER) unter: <http://www.aeaweb.org/aer/data.php>



Software

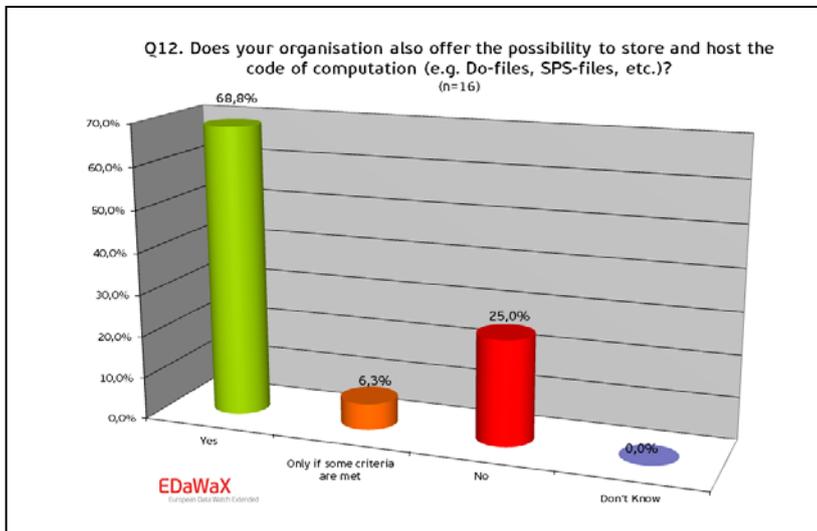
In Bezug auf Speicherung und Hosting von (selbstgeschriebener) Software, wie sie etwa im Rahmen wirtschaftswissenschaftlicher Simulationen Verwendung findet, zeigte sich, dass nur eine Minderheit von knapp einem Viertel der untersuchten Organisationen eine Speicherung und das Hosting solcher Daten ohne 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 ist). 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.



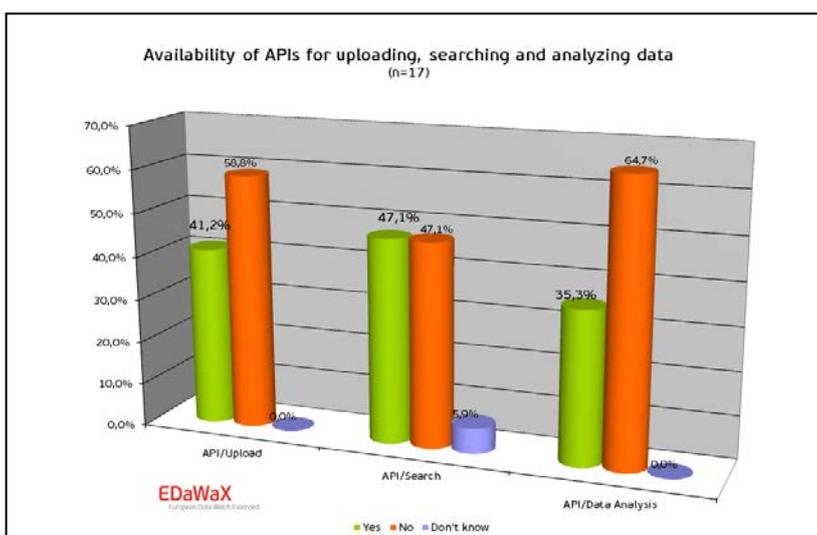
Syntax

Fast 70% der untersuchten Organisationen bieten die Möglichkeit Syntax zu speichern und bereit zu stellen – ¼ 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.



APIs

Im Zuge unserer Untersuchung wurde auch die Verfügbarkeit von Programmierschnittstellen (so genannten 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.



Unsere Befragung ergab, dass weniger als die Hälfte aller Organisationen angaben, über solche Schnittstellen zu verfügen. Am häufigsten wurden APIs zur Suche von Datensätzen genannt (47%), gefolgt von APIs für den Upload von

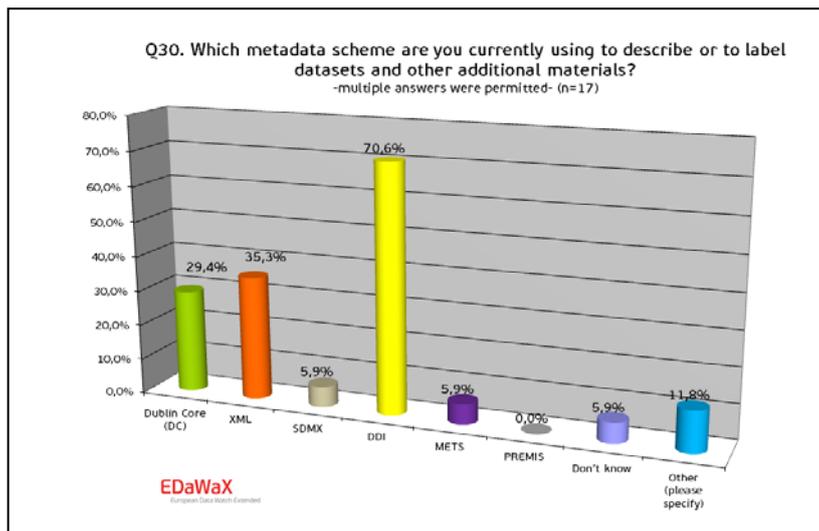
Forschungsdaten. Etwas mehr als ein Drittel (35%) der Befragten gibt zudem an, über Schnittstellen für die Analyse von Forschungsdaten zu verfügen.

Eine vom EDaWaX-Projekt 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

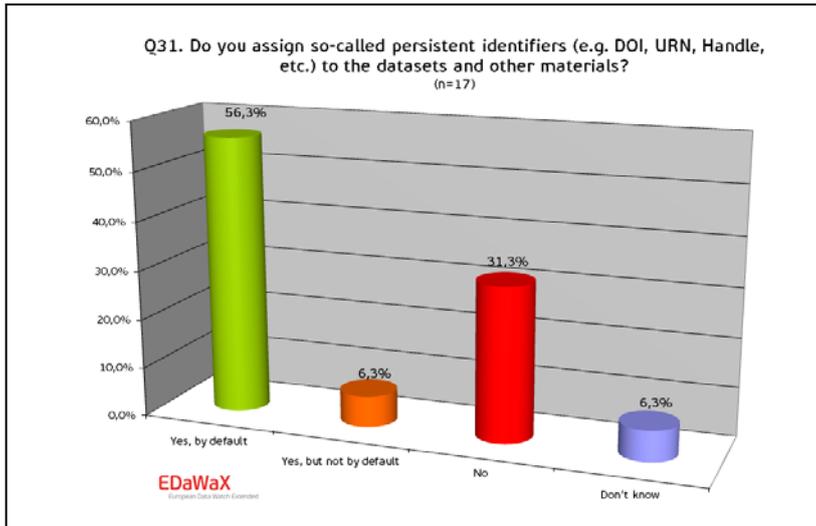
Genutzte Metadatenschemata

Das Projekt interessierte sich zudem für die Metadatenschemata, die die befragten Organisationen bei ihrer Arbeit verwenden. Dabei zeigte sich, dass mehr als 70% der Befragten DDI verwenden. Wesentlich seltener wurde XML (35%) oder DC (knapp 30%) genannt. Alle anderen Metadatenschemata werden nur vereinzelt genutzt.



Persistente Identifier (PI)

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

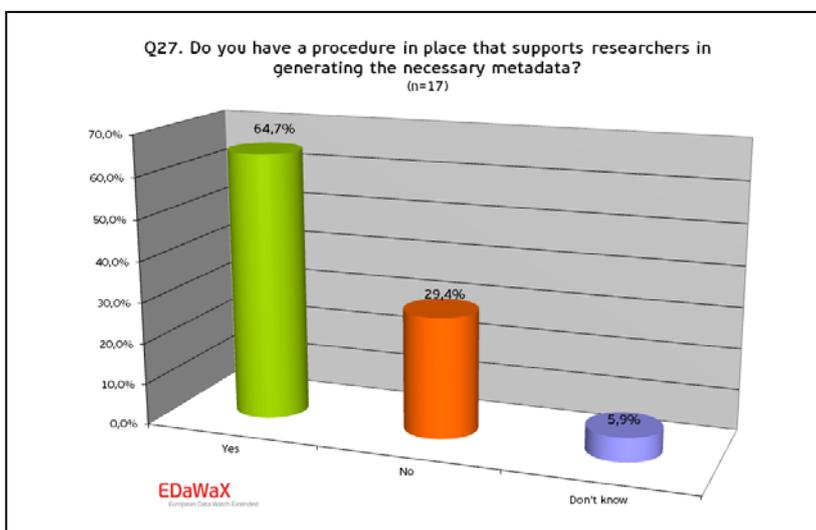


Unterstützung von Semantic Web Technologien

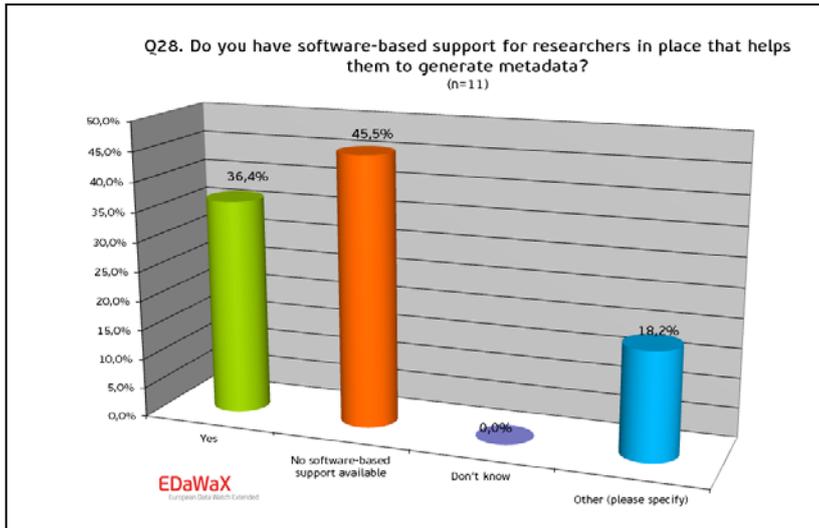
In unserer Befragung wurde auch nach der Verwendung von RDF gefragt – einem Datenmodell zur Beschreibung von Ressourcen mittels semantischer Technologien. Von den befragten Organisationen gab nur eine Minderheit von etwa 6% an, RDF-Daten bereit zu stellen. Fast ¼ der Befragten machte hierzu keine Angaben, was teilweise auf einen geringen Bekanntheitsgrad von RDF zurückzuführen sein dürfte.

Support bei der Metadatenerstellung

Die Achillesferse für die Nutzbarkeit 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.



Unsere Befragung ergab, dass die Mehrheit der Organisationen (fast 65%) über einen entsprechenden Support für Wissenschaftler/innen verfügt.

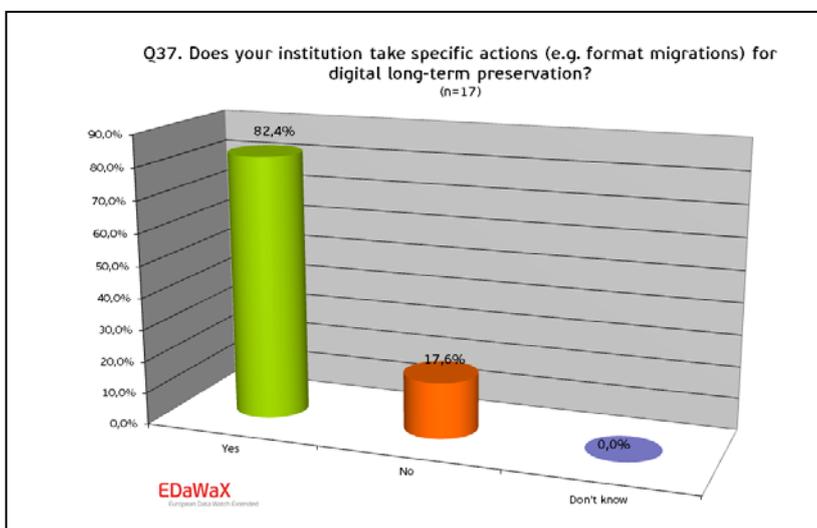


Dabei interessierte uns auch, ob es softwarebasierte Unterstützung für die Erstellung von Metadaten an den Institutionen gibt, wie dies beispielsweise durch entsprechende Eingabemasken und die Konvertierung der eingegebenen Inhalte in standardisierte Metadaten der Fall sein kann. Hier zeigte sich, dass über 35% der Befragten über eine solchen softwarebasierten Support verfügen. Auffällig ist die Zahl an Nennungen im Bereich *other*. Hier wurden beispielsweise auch schriftlich auszufüllende *Data Deposit Forms* aufgeführt.

Unsere Nachfrage nach dem Namen bzw. der Art der Software ergab, dass mindestens zwei Institutionen Nesstar¹³ einsetzen. Viele Organisationen nutzen zudem Eigenentwicklungen.

Langzeitarchivierung

In unserer Befragung wollten wir auch wissen, in welchem Umfang Maßnahmen zur Langzeitarchivierung von Forschungsdaten getroffen werden. Hier zeigte sich, dass mehr als 80% der Befragten entsprechende Maßnahmen umsetzen.



Fazit

Die Befragungsergebnisse zeigen, dass Forschungsdatenzentren ein relevanter Speicherort für publikationsbezogene Forschungsdaten sein können, da sie verschiedene Voraussetzungen dafür bereits erfüllen. Dennoch gibt es unter den befragten Organisationen bislang keine Institution die in Gänze alle Anforderungen hinsichtlich Speicherung und Hosting solcher publikationsbezogenen Forschungsdaten in den Wirtschaftswissenschaften erfüllt.

Im Einzelnen ergaben sich folgende Ergebnisse:

- Etwa Dreiviertel aller befragten Einrichtungen akzeptieren grundsätzlich externe Forschungsdaten, 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 Befragten 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 Dreiviertel 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 befragten Organisationen sorgen zudem zu über 80% für die Langzeitverfügbarkeit der bei ihnen gehosteten Datensätze.

Literaturverzeichnis

- 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. / Ringnalda, A. (2011): The legal status of research data in the Knowledge Exchange partner countries. Verfügbar unter: <http://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: http://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/univlag/2012/GSI8_Hillegeist.pdf
- King, G. (1995). Replication, replication. In: *PS: Political Science and Politics*, 28, 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), 117-126
- Research Information Network (2011). Data centres: their use, value and impact. A Research Information Network report. September 2011. Verfügbar unter: http://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
- Siegert, 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 Verlag.
- Vlaeminck, S. (2013). “Data Management in Scholarly Journals and possible Roles for Libraries – Some Insights from EDaWaX.” *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. / Siegert, O. (2012). “Welche Rolle spielen Forschungsdaten eigentlich für Fachzeitschriften? Eine Analyse mit Fokus auf die Wirtschaftswissenschaften.” *RatSWD Working Papers*, Nr. 210, Berlin. Verfügbar unter: http://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) (auch erschienen als RatSWD Working Paper Nr. 224)