

Vlaeminck, Sven; Siegert, Olaf

**Conference Paper**

## Welche Rolle spielen Forschungsdaten eigentlich für Fachzeitschriften? Eine Analyse mit Fokus auf die Wirtschaftswissenschaften

*Suggested Citation:* Vlaeminck, Sven; Siegert, Olaf (2012) : Welche Rolle spielen Forschungsdaten eigentlich für Fachzeitschriften? Eine Analyse mit Fokus auf die Wirtschaftswissenschaften, In: Vernetztes Wissen - Daten, Menschen, Systeme, ISBN 978-389-33682-1-1, Forschungszentrum Jülich, Jülich, pp. 29-44, <http://hdl.handle.net/2128/4699>

This Version is available at:  
<http://hdl.handle.net/11108/80>

**Kontakt/Contact**

ZBW – Leibniz-Informationszentrum Wirtschaft/Leibniz Information Centre for Economics  
Düsternbrooker Weg 120  
24105 Kiel (Germany)  
E-Mail: [info@zbw.eu](mailto:info@zbw.eu)  
<http://zbw.eu/de/ueber-uns/profil/veroeffentlichungen-zbw/>

**Standard-Nutzungsbedingungen:**

Dieses Dokument darf zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden. Sie dürfen dieses Dokument nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Sofern für das Dokument eine Open-Content-Lizenz verwendet wurde, so gelten abweichend von diesen Nutzungsbedingungen die in der Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*This document may be saved and copied for your personal and scholarly purposes. You are not to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public. If the document is made available under a Creative Commons Licence you may exercise further usage rights as specified in the licence.*

## **Welche Rolle spielen Forschungsdaten eigentlich für Fachzeitschriften? Eine Analyse mit Fokus auf die Wirtschaftswissenschaften**

Sven Vlaeminck, Olaf Siegert

**ZBW - Leibniz Informationszentrum Wirtschaft**

### **Zusammenfassung**

In den Wirtschafts- und Sozialwissenschaften nehmen empirische Untersuchungen an Bedeutung zu. Dies schlägt sich auch in einer steigenden Anzahl von entsprechenden Beiträgen in Fachzeitschriften nieder, bei denen Autoren<sup>1</sup> eigene Forschungsdaten erhoben oder externe Datensätze verarbeitet haben. Allerdings bieten bislang nur wenige Zeitschriften effektive Möglichkeiten, die verwendeten Forschungsdaten im Kontext der zugehörigen Zeitschriftenartikel aufzurufen. Auch eine Zitation, Replikation oder eine Nachnutzung dieser Daten ist oftmals nicht möglich. Unser Paper beschreibt zunächst den Status Quo in Bezug auf das Forschungsdatenmanagement in wirtschaftswissenschaftlichen Fachzeitschriften. In einem zweiten Schritt werden die Ergebnisse einer Studie des DFG-geförderten Projekts EDaWaX (European Data Watch Extended) präsentiert und diskutiert, bei der die Richtlinien von Fachzeitschriften zur Übermittlung von Forschungsdaten untersucht wurden. Aufbauend auf den Ergebnissen der Studie wird die Frage aufgeworfen, wie eine sinnvolle Bereitstellung solcher publikationsbezogenen Forschungsdaten erfolgen kann.

### **Abstract**

Empirical studies to verify and refine theoretical models are increasingly important in economics. This is reflected in a rising number of empirical contributions to journals where the authors have collected their own research data or used external datasets. However, so far there have been few effective means to access the results of economic research within the framework of the corresponding article. Also to cite, to replicate or to make the research data that was used in a paper available for repurposing often is not possible. Initially our paper describes the status quo in regard to the management of research data in economic scholarly journals. In a second step the results of an empirical study conducted by the DFG-funded research project EDaWaX (European Data Watch Extended) are presented and discussed. The project evaluated the data availability policies of some economic scholarly journals. Based on the results of the study the question is raised how publication-related research data can reasonably be made available.

---

<sup>1</sup> Die in dieser Publikation gewählte männliche Form bezieht immer gleichermaßen alle Geschlechter mit ein. Auf eine Mehrfachbezeichnung wurde nur aufgrund einfacherer Lesbarkeit verzichtet.

### **Einleitung**

In den Wirtschaftswissenschaften existieren gegenwärtig schätzungsweise rund 8.000 bis 10.000 Fachzeitschriften<sup>2</sup>, davon werden nach heutigem Stand fast 500 Open Access bereitgestellt<sup>3</sup>. Der wissenschaftliche Zeitschriftenmarkt in den Wirtschaftswissenschaften wird daher nach wie vor von klassischen Subskriptionszeitschriften dominiert.

Längst nicht alle dieser vielen tausend Zeitschriften gehören zu den relevanten Publikationen der Fachdisziplin: So listet Thomson Reuters Journal Citation Reports (JCR)<sup>4</sup> im Jahr 2011 rund 320 wirtschaftswissenschaftliche Fachzeitschriften auf, die über einen Impact Factor verfügen. Das renommierte Handelsblatt Ranking vergab für das Jahr 2010 einen Impact Factor für 1250 Zeitschriften aus dem Bereich Volkswirtschaftslehre<sup>5</sup> und 750 Zeitschriften aus dem Bereich Betriebswirtschaftslehre<sup>6</sup>. Diese Zahlen zeigen, dass die Zahl der relevanten Zeitschriften noch einigermaßen übersichtlich ist.

Nicht alle diese Zeitschriften veröffentlichen empirische Beiträge. So gibt es eine nicht unerhebliche Anzahl von Zeitschriften die ausschließlich theoretische Arbeiten publizieren, oder Artikel deren Fokus auf Politikberatung liegt.

Insgesamt jedoch nimmt die Zahl an Veröffentlichungen in wirtschaftswissenschaftlichen Fachzeitschriften zu, bei denen Forscher eigene Datensätze erhoben und ausgewertet haben, oder bereits bestehende Datensätze für ihre Fragestellungen verwendet haben. Effektive Möglichkeiten die verwendeten Forschungsdaten im Kontext der zugehörigen Zeitschriftenartikel aufzurufen oder zu zitieren, bieten bislang allerdings nur wenige Journals. Auch eine Prüfung oder Nachnutzung dieser Daten zur Unterstützung des wissenschaftlichen Diskurses ist oftmals nicht möglich. Dabei ist eine Prüfung und Replikation der postulierten Forschungsergebnisse nicht nur vor dem Hintergrund zahlreicher Fälle von wissenschaftlichem Fehlverhalten<sup>7</sup> geboten, sondern es sprechen, wie

---

<sup>2</sup> [http://open-access.net/de/oa\\_in\\_verschiedenen\\_faechern/wirtschaftswissenschaften](http://open-access.net/de/oa_in_verschiedenen_faechern/wirtschaftswissenschaften)  
(Aufruf: 15. August 2012)

<sup>3</sup> Vgl. Directory of Open Access Journals (DOAJ):  
<http://www.doaj.org/doi?func=subject&cpid=17&uiLanguage=en> (Aufruf: 15. August 2012)

<sup>4</sup> [http://thomsonreuters.com/products\\_services/science/science\\_products/a-z/journal\\_citation\\_reports/](http://thomsonreuters.com/products_services/science/science_products/a-z/journal_citation_reports/)

<sup>5</sup> Handelsblatt Ranking VWL: <http://tool.handelsblatt.com/tabelle/?id=33> (Aufruf: 15. August 2012)

<sup>6</sup> Handelsblatt Ranking BWL: <http://tool.handelsblatt.com/tabelle/?id=34&so=1a&pc=25&po=0> (Aufruf: 15. August 2012)

<sup>7</sup> Ein bekanntes Beispiel für wissenschaftliches Fehlverhalten stellt der niederländische Sozialpsychologe Diederik Stapel, der zahlreiche seiner Forschungsdaten vorsätzlich gefälscht hat. Vgl. <http://www.nature.com/news/2011/111101/full/479015a.html>

beispielsweise der US-Ökonom B.D. McCullough argumentiert, verschiedene Gründe für die Replizierbarkeit von Forschungsdaten: „[...] *replication ensures that the method used to produce the results is known. Whether the results are correct or not is another matter, but unless everyone knows how the results were produced, their correctness cannot be assessed. Replicable research is subject to the scientific principle of verification; non-replicable research cannot be verified. Second, and more importantly, replicable research speeds scientific progress. [...] Third, researchers will have an incentive to avoid sloppiness. [...] Fourth, the incidence of fraud will decrease (McCullough, 2009, 118f).*“

Dabei könnte von technologischer Seite heutzutage der Zugriff auf Forschungsdaten so leicht wie nie zuvor sei: Das Internet hat die vormals dominierenden Printpublikationen nun zunehmend mit Onlineausgaben der Zeitschriften ergänzt oder diese sogar ersetzt. Dadurch ist prinzipiell der Zugriff auf in Artikeln genutzte Forschungsdaten möglich und sogar deren weitere Bearbeitung und Nachnutzung. Damit solche Forschungsdaten aber verfügbar sind, müssen die Journals einerseits Richtlinien aufstellen, mit denen diese Daten eingefordert werden. Andererseits muss eine technische Infrastruktur für die Speicherung und das Auffinden von Forschungsdaten im Web bereitstehen.

Forschungsdatenrichtlinien manifestieren sich zumeist in so genannten Data (Availability) Policies, die Autoren dazu verpflichten, die genutzten Forschungsdaten zusätzlich zur Publikation bei den Herausgebern einer Fachzeitschrift einzureichen.<sup>8</sup> Gegenwärtig verfügt nur ein sehr geringer Anteil von Zeitschriften aus unterschiedlichen Gründen über solche Richtlinien. So gibt es auf Seiten der Herausgeber Befürchtungen, dass sich die Einführung einer solchen Policy negativ auswirken könnte: Geringere Einreichungsquoten werden hierbei ebenso genannt, wie potenziell zusätzlicher Aufwand durch komplexere Kommunikation mit Autoren einerseits, sowie durch Sichtung der Daten und eine Prüfung, ob die Vorgaben der Data Policy eingehalten werden andererseits.

---

<sup>8</sup> In der wissenschaftlichen Literatur wurde dieser Themenbereich vor allem durch B.D. McCullough et al (2006, 2007, 2008, 2009) bearbeitet. Weitere relevante Veröffentlichungen erschienen durch die Ökonomen Dewald, Thursby und Anderson (1986), sowie durch Anderson et al (2008) und Glandon (2010).

Auch der infrastrukturelle Aufwand durch eine Bereitstellung der übermittelten Daten im Internet wird als Hinderungsgrund angesehen.<sup>9</sup>

Infolgedessen konnte McCullough im Rahmen einer 2009 veröffentlichten Studie gerade einmal 14 solcher Zeitschriften in den Wirtschaftswissenschaften ausmachen, was ihn zu deutlicher Kritik veranlasste: "Results published in economic journals are accepted at face value and rarely subjected to the independent verification that is the cornerstone of the scientific method. Most results published in economics journals cannot be subjected to verification, even in principle, because authors typically are not required to make their data and code available for verification (McCullough/McGeary/Harrison, 2006, 1093f.)."

#### **Data Policies in wirtschaftswissenschaftlichen Fachzeitschriften**

Den aktuellen Status Quo in Bezug auf das Vorhandensein von Data Policies in wirtschaftswissenschaftlichen Fachzeitschriften zu evaluieren, war Ziel einer empirischen Untersuchung des DFG-geförderten Projekts „European Data Watch Extended“ (EDaWaX). In der Studie wurde untersucht, wie sich der Anteil von Zeitschriften mit entsprechenden Richtlinien in den letzten Jahren verändert hat, ob Zeitschriften mit Data Policies schwerpunktmäßig in bestimmten Verlagen oder Verlagsgruppen publiziert werden und wie Journals mit solchen Richtlinien in verschiedenen wirtschaftswissenschaftlichen Zeitschriftenrankings bewertet wurden. Ausführlich wurde ausgewertet, wie die Data Availability Policies der Zeitschriften jeweils im Detail ausgestaltet sind, welche konkreten Vorgaben den Autoren darin gemacht werden und in wie fern die Richtlinien eine Replizierbarkeit der publizierten Forschungsergebnisse ermöglichen.

#### Sample und Methodik

Für die Untersuchung wurde eine Sample aus insgesamt 141 Fachzeitschriften gebildet, das sich an einer Liste von Fachzeitschriften orientiert, die von den Ökonomen Bräuninger, Haucap und Muck (im Folgenden: BHM) im Jahre 2011 erstellt wurde. Diese Zeitschriften wurden seinerzeit durch Mitglieder der wirtschaftswissenschaftlichen Fachgesellschaft Verein für Socialpolitik (VfS)<sup>10</sup> hinsichtlich ihrer Relevanz und Reputation bewertet: „Die ausgewählten Zeitschriften

---

<sup>9</sup> Diese Einschätzungen werden auch von einer Onlinebefragung von fast 200 wirtschaftswissenschaftlichen Zeitschriften gestützt, die von den Autoren im Rahmen des EDaWaX-Projekts durchgeführt wurden. Eine Summary der Ergebnisse dieser Befragung wird im Herbst 2012 auf der Projektwebseite [www.edawax.de](http://www.edawax.de) publiziert.

<sup>10</sup> <http://www.socialpolitik.org/vfs.php?mode=start>

umfassen die wichtigsten in internationalen Ranglisten berücksichtigten Fachzeitschriften sowie einen Großteil der deutschsprachigen bzw. in Deutschland, Österreich oder der Schweiz herausgegebenen deutsch- und englischsprachigen VWL-Fachzeitschriften [...] (BHM, 2011, 2)“. Dieses Sample bot verschiedene Vorteile für die Auswertung: So ließen sich Zeitschriften aus dem deutschen Sprachraum mit den internationalen Top Journals vergleichen, und weitere Analysen in Bezug auf Relevanz und Reputation der Zeitschriften mit Data Availability Policy vornehmen.

Ergänzt durch vier Zeitschriften mit Data Availability Policies wurde insgesamt ein Sample genutzt, in dem solche Zeitschriften deutlich stärker vertreten sind, als dies durchschnittlich im Bereich der Wirtschaftswissenschaften der Fall sein dürfte, was jedoch aufgrund des Untersuchungsgegenstandes vertretbar ist. Zudem ist das Untersuchungssample durch Einbeziehung vieler Top-Journals überdurchschnittlich hoch gerankt.

Die 141 Zeitschriften unseres Samples wurden zunächst mittels Desktop-Research untersucht. Dabei wurden sowohl die Webseiten des Verlages als auch die der Herausgeber nach Richtlinien untersucht, die den Umgang mit Forschungsdaten regeln.<sup>11</sup> Für eine deskriptive Auswertung des Rankings der untersuchten Zeitschriften wurden Mittelwerte und Median für das Sample erstellt. Zudem wurden logistische Regressionen durchgeführt, um mögliche Zusammenhänge zwischen dem erzielten Ranking und dem Vorhandensein von Data Policies zu prüfen.

Daran anschließend wurden die Inhalte der aufgefundenen Data Policies anhand von Kriterien analysiert, die in der Literatur als wesentliche Elemente zur Ermöglichung von Replikationen benannt und durch eigene Überlegungen des Projekts ergänzt wurden.

Darüber hinaus wurden eine Onlinebefragung von Herausgebern von fast 200 Fachzeitschriften sowie zwei detaillierte Case Studies durchgeführt.

---

<sup>11</sup> Die Printausgaben der Zeitschriften wurden nicht geprüft.

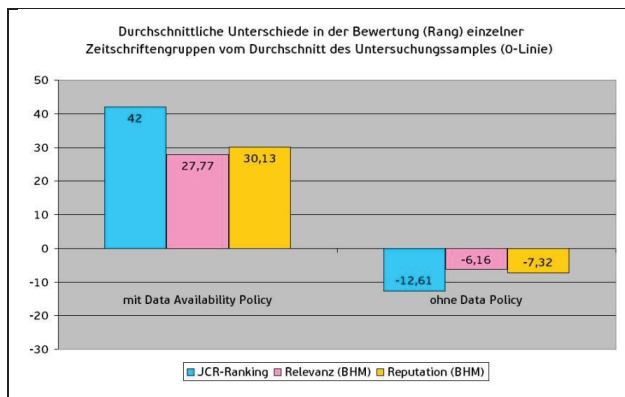
### Deskriptive Ergebnisse

Mehr als 20% (29) der untersuchten 141 Zeitschriften verfügen über eine Data Policy, bei der Daten und weitere Materialien an die Herausgeber übermittelt werden müssen, die diese wiederum zumeist für Dritte bereitstellen. Sie entsprechen somit häufig dem von Harvard Professor Gary King formulierten „Replication Standard“ (King, 1995). Der relativ hohe Wert von über 20% ist aufgrund unserer Sampleauswahl aber keinesfalls als repräsentativ für die Wirtschaftswissenschaften anzusehen.

Zeitschriften mit solchen Richtlinien wurden am häufigsten durch die Verlage Wiley-Blackwell (6), und Elsevier (4) publiziert. Im Vergleich zeigte sich jedoch, dass anteilig vor allem Verlage aus dem universitären Kontext (z.B. Cambridge University Press) bzw. von wissenschaftlichen Fachgesellschaften (z.B. American Economic Association) über hohe bis sehr hohe Anteile an Zeitschriften mit Data Availability Policies verfügen.

Zudem konnten wir feststellen, dass Zeitschriften mit Data Availability Policy deutlich besser gerankt waren als Zeitschriften ohne solche Richtlinien:

Der Impact Factor dieser Zeitschriften im JCR liegt durchschnittlich fast 0,5 Punkte höher als der Mittelwert unseres Samples. Beim



Handelsblatt Ranking waren es durchschnittlich 0,21 Punkte mehr.

Ähnlich deutlich fallen die Ergebnisse bei der Platzierung der Zeitschriften mit Data Availability Policy im Vergleich zu solchen ohne

derartige Richtlinien aus: So wurden im JCR Zeitschriften mit Data Availability Policy durchschnittlich 42 Plätze höher bewertet als der Mittelwert. Bei den Rankings von BHM waren es durchschnittlich 27,8 Plätze (bei Relevanz) bzw. 30,1 Plätze (bei Reputation). Die durchgeführten Regressionsanalysen unterstützen diese Angaben, da für alle Rankings ein signifikanter mittlerer Zusammenhang zwischen einem

hohen Ranking und dem Vorhandensein einer Data Availability Policy ermittelt wurde.<sup>12</sup>

Somit ist festhalten, dass es vor allem die Top-gerankten Zeitschriften sind, die Data Availability Policies umsetzen. Dies mag einerseits mit den notwendigen zusätzlichen Ressourcen in Verbindung stehen, die für die organisatorische und technische Umsetzung einer solchen Richtlinie nötig sind. Andererseits kann vermutet werden, dass es gerade Publikationen in solchen prestigeträchtigen, hochgerankten Journals sind, die Wissenschaftlern die Anreize verschaffen, die oft umfangreiche und zeitaufwändige Dokumentation und Bereitstellung von Forschungsdaten vorzunehmen – eine Arbeit für die das Wissenschaftssystem bislang sonst keine angemessene Anerkennung liefert (Huschka et al, 2011).

#### Kriterien für die inhaltliche Untersuchung von Data Policies

Die Inhalte der aufgefundenen Data Availability Policies wurden in einem zweiten Schritt anhand von 9 inhaltlichen Kriterien analysiert, die in der Literatur als Wesentlich für das Gelingen von Replikationen charakterisiert werden und durch eigene Überlegungen ergänzt wurden. Folgende fachspezifische Empfehlungen für Data Policies wurden entwickelt:

- Die Richtlinien sollten verbindlich sein.
- Autoren sollten die verwendeten Datensätze an die Herausgeber übermitteln.
- Autoren sollten die verwendeten Berechnungscodes und Skripte an die Herausgeber übermitteln.
- Autoren sollten ausführliche Beschreibungen der übermittelten Daten sowie der Variablen und Merkmalsausprägungen der verwendeten Datensätze, am besten in Form eines „Data Dictionary“, an die Herausgeber übersenden.
- Wurden bei den Berechnungen (selbstgeschriebene) Programme verwendet (z.B. im Rahmen von Simulationen), sollten Autoren diese verwendeten Programme an die Herausgeber übermitteln.
- Die Datensätze und weitere Materialien sollten bereits vor der Publikation eines Artikels an die Herausgeber übermittelt werden.
- Wenn Ausnahmen von den Richtlinien (z.B. für Forschungen auf Basis von proprietären Daten) gestattet werden, sollte in den Richtlinien ein Vorgehen definiert sein, welche Daten dennoch an die Herausgeber übermittelt werden sollten (z.B. ausführliche Beschreibungen und Berechnungscodes).

---

<sup>12</sup> Zwar erklärt die Data Availability Policy nur zu 8,7% bis 16% (je nach untersuchtem Ranking) das hohe Ranking von Zeitschriften mit einer Data Availability Policy, dies jedoch auf einem Signifikanzniveau zwischen 99,7% und 100%.



- Datensätze und weiteren Materialien sollten aus Gründen der Interoperabilität sowie der Langzeitarchivierung im ASCII-Format oder in anderen offenen Formaten an die Herausgeber übermittelt werden.
- Die Version der für die Berechnungen genutzten Software sowie des verwendeten Betriebssystems sollte bei der Übermittlung der Daten mit angegeben werden.

Unsere Prüfung der aufgefundenen Data Policies auf die oben genannten Empfehlungen ergab folgende Ergebnisse:

**Verbindlichkeit der Data Policies:** Wir konnten feststellen, dass 82,8% der Data Availability Policies für Autoren verpflichtend waren. Diese hohen Prozentsätze sind begrüßenswert, da vor allem verpflichtende Policies von den Autoren befolgt werden (McCullough et al., 2008).

**Übermittlung von Datensätzen:** Das Vorhandensein der genutzten Datensätze ist die erste Bedingung zur Durchführung von Replikationen. Fast 90% der untersuchten Data Policies sahen eine verpflichtende Übermittlung von Datensätzen vor. Unter den 10% der Zeitschriften, die keine verbindliche Übermittlung von Datensätzen vorsahen, waren einige, deren Fokus vor allem auf experimenteller Wirtschaftsforschung lag. Bei diesen ist die Übermittlung von Datensätzen zumeist optional.



**Übermittlung von Berechnungscode:** Replikationen sind ohne den passenden Berechnungscode sehr mühselig, da Forschende diesen erneut schreiben müssen. Ob dieser dann mit dem Berechnungscode des Autors übereinstimmt, ist

fraglich. Daher ermöglichen Zeitschriften, die ausschließlich die verwendeten Datensätze einfordern, nicht aber den dazu passenden Code, oftmals keine Replikationen (Dewald/Thursby/Anderson, 1986).

Unsere Untersuchung ergab, dass nur 51,7% der untersuchten Data Availability Policies die Übermittlung dieses Berechnungscodes einforderten. An dieser Stelle zeigen sich deutlich die qualitativen Probleme vieler untersuchten Richtlinien.

**Übermittlung von Beschreibungen:** Nur 65% der untersuchten Data Availability Policies forderten Beschreibungen zu den übermittelten Daten mit ein. Dieser Wert ist nicht zufriedenstellend, da eine Beschreibung der übermittelten Daten notwendig ist, um zu verstehen, welche der übermittelten Daten mit welchen Ergebnissen in einem Artikel korrespondieren (McCullough, 2007). Darüber hinaus ist die Beschreibung von Datensätzen, Variablen und Merkmalsausprägungen für die Durchführung von Replikationen sowie für die Nachvollziehbarkeit der Ergebnisse von hoher Relevanz.

**Übermittlung von Programmen:** 62% der untersuchten Data Availability Policies forderten die Übermittlung genutzter, meist selbst geschriebener Programme, wie sie häufig im Rahmen von Simulationen verwendet werden. Auch dieser Wert ist nicht zufriedenstellend, da es ohne die genutzten Programme oft nicht möglich ist, zu prüfen, ob die Ergebnisse korrekt berechnet wurden. Folglich können Ergebnisse dann nicht repliziert werden (McCullough/McGeary/Harrison, 2006).

**Zeitpunkt der Datenübermittlung:** Mehr als 90% der untersuchten Data Availability Policies sehen vor, dass Daten und Datensätze bereits vor der Veröffentlichung der Forschungspublikation an die Herausgeber übermittelt werden müssen. Hier haben die Herausgeber offensichtlich aus den Erfahrungen anderer Data Policies gelernt, die zeigen, dass eine verbindliche Übermittlung der Daten vor allem dann erfolgt, wenn es für die Wissenschaftler einen Anreiz gibt, die Daten zu publizieren. Nach der Publikation eines Artikels ist dieser Anreiz nicht mehr in dieser Form gegeben (Dewald/Thursby/Anderson, 1986).

**Definiertes Vorgehen im Fall proprietärer Daten:** Ausnahmen von den Richtlinien im Fall der Nutzung von proprietären oder vertraulichen Daten sehen mindestens 72,4% der Data Availability Policies vor, aber nur rund 60% verfügen über ein Verfahren, das regelt, welche Daten die Autoren dennoch an die Herausgeber einer Zeitschrift übermitteln müssen. Auch bei Nutzung proprietärer oder vertraulicher Daten sollten Autoren weiterhin dazu verpflichtet sein, Berechnungscodes und

Beschreibungen zu den Daten weiter zu geben, so dass Replikationen im Prinzip auch bei solchen Forschungsdaten möglich sind.

**Formatvorgaben:** Vorgaben zur Nutzung bestimmter Formate, wie ASCII oder von offenen Formaten werden nur in Ausnahmefällen gemacht. Nur einzelne Zeitschriften gehen auf diese Problematik ein, und halten ihre Autoren an, die Daten (auch) in solchen Formaten bereitzustellen. Dabei sind offene Formate bzw. ASCII besonders unter den Gesichtspunkten der interoperablen Nutzung von Daten und Code, aber auch in Hinsicht auf eine Langzeitarchivierung der Daten sinnvoll (McCullough/McGeary/Harrison, 2006).

**Angabe der Version der genutzten Software und des Betriebssystems:** Die Angabe der genutzten Versionen der Software und des Betriebssystems ist wichtig, weil die erzielten Forschungsergebnisse nach genutzter Software und verwendetem Betriebssystem divergieren können (McCullough/Vinod, 2003).

Bis heute sieht keine der untersuchten Richtlinien eine Angabe des bei den Berechnungen verwendeten Betriebssystems vor und nur rund ein Drittel der Data Availability Policies fordert zudem die Angabe der genutzten Softwareversion.

### **Ergebnisse**

Zusammenfassend ist festzustellen, dass sich die Anzahl von Zeitschriften mit Data Availability Policy seit den letzten Forschungen von McCullough (2009) deutlich erhöht hat. Es ist somit ein gesteigertes Interesse der Herausgeber an solchen Richtlinien erkennbar.

Unsere Untersuchung hat ergeben, dass die analysierten Data Availability Policies teils erheblich in Umfang, Qualität und in der Granularität, in der Vorgaben für Autoren gemacht werden, divergieren. Dabei zeigte sich, dass die untersuchten Richtlinien nicht in allen Punkten inhaltlich gut ausgestaltet sind: So sind zwar rund 90% der Richtlinien verbindlich und sogar mehr als 90% verlangen die Übermittlung von Daten und Datensätzen bereits vor der Publikation des zugehörigen Artikels, jedoch fordern nur  $\frac{3}{4}$  der untersuchten Policies genutzte Programme oder Beschreibungen zu den Daten ein. Auch gewähren mehr als 70% der untersuchten Data Policies Ausnahmen von ihren Forschungsdatenrichtlinien im Fall der Verwendung proprietärer oder vertraulicher Daten, aber nur rund 60% haben

definiert, welche Daten und Angaben Autoren in solchen Fällen trotzdem übermitteln müssen. Hier zeigen sich deutliche inhaltliche Schwächen bei einigen Data Policies. Als problematisch muss insbesondere gelten, dass nur etwas mehr als die Hälfte der Richtlinien zugleich den Berechnungscode mit einfordert - für die angestrebte Replizierbarkeit von Forschungsergebnissen dürften solche Policies nicht ausreichen. Auch das weitgehende Fehlen von Metadaten zu den Versionen der Software und des Betriebssystems, welche bei der Berechnung der Ergebnisse genutzt wurde, ist kritikwürdig. Der weitgehende Verzicht auf Formatvorgaben ist speziell in Hinsicht auf eine interoperable Nutzung der Daten und die Langzeitverfügbarkeit von Forschungsdaten problematisch.

Aufgrund der erzielten Ergebnisse ist zu vermuten, dass nur etwa die Hälfte der untersuchten Data Availability Policies Replikationen gewährleisten dürfte.

Besonders gut ausgearbeitet war die Data Availability Policy der American Economic Review (AER)<sup>13</sup>. Dieser Richtlinie kann als Best Practice unter den aufgefundenen Data Policies bezeichnet werden. Damit, dass 34,5% der untersuchten Data Policies diese Richtlinie in identischer oder leicht modifizierter Form nutzen, ist eine positive Entwicklung in Richtung einer sich etablierenden Best-Practice in den Wirtschaftswissenschaften zu beobachten.

### **Infrastruktur zur Bereitstellung von publikationsbezogenen Forschungsdaten**

Das Vorhandensein von Data Availability Policies ist nur eine Bedingung für die Nachnutzung von Forschungsdaten oder zur Replikation der Forschungsergebnisse. Eine andere Bedingung ist, dass eine entsprechende technische Infrastruktur zur Bereitstellung dieser publikationsbezogenen Forschungsdaten zur Verfügung steht.

In der wissenschaftlichen Forschung wird dieses Thema vor allem von Institutionen bearbeitet, die die Bereitstellung von Forschungsdaten übernehmen (z.B. Wagner und Huschka, 2012; Huschka et al, 2011) als auch von übergeordneten Kommissionen und Projekten<sup>14</sup>, die teilweise auch dezidierte Vorschläge zur Entwicklung einer Forschungsdateninfrastruktur für publikationsbezogene Forschungsdaten enthalten.

---

<sup>13</sup> <http://www.aeaweb.org/aer/data.php/>

<sup>14</sup> Vgl. Allianz der deutschen Wissenschaftsorganisationen, 2010; Reilly et al, 2011; Kommission zur Zukunft der Informationsinfrastruktur, 2011; Wissenschaftsrat, 2011; Wissenschaftsrat, 2012

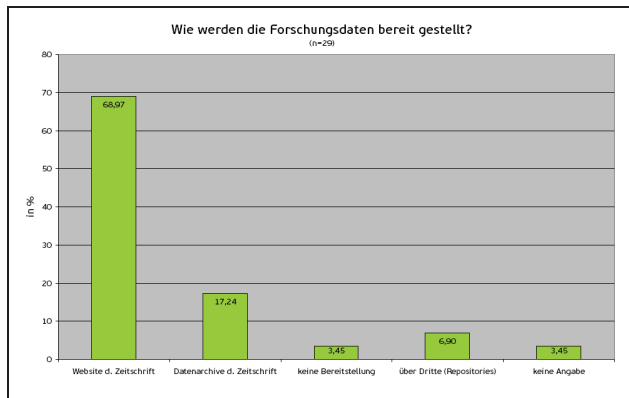
Auf Seiten der notwendigen technischen Infrastruktur für die Etablierung von zitierfähigen Forschungsdatenpublikationen gibt es sowohl in der Bundesrepublik als auch international übergreifende Ansätze, auf denen aufgebaut werden kann. So wurde im Jahr 2001 mit der Einrichtung der ersten vier Forschungsdatenzentren (FDZ) und zwei Datenservicezentren (DSZ) begonnen.<sup>15</sup> Ziel dieser Datenzentren ist es, die amtlichen Daten zur wissenschaftlichen Nutzung zur Verfügung zu stellen. Dies war bis dahin aufgrund von gesetzlichen Vorgaben nicht ohne weiteres möglich (Huschka et al, 2011). Neben diesen existieren weltweit unterschiedliche Datenarchive mit wirtschafts- und sozialwissenschaftlich relevanten Daten (z. B. Mikrodaten zu privaten Haushalten und Unternehmen oder Makrodaten zum Arbeitsmarkt oder zum Außenhandel). Diese Datenarchive verfügen über Erfahrungen bei der Beschreibung von Datensätzen und haben unter anderem den internationalen Metadatenstandard DDI<sup>16</sup> (Data Documentation Initiative) entwickelt. Ebenso verfügen viele Datenarchive über langjährige Erfahrungen im Bereich der Zugangsregelungen und Datenschutzbestimmungen bezüglich ihrer Daten (Reilly et al, 2011) sowie in Bezug auf den Langzeiterhalt dieser wichtigen Quellen von Forschung und Lehre (Mauer, 2011). Speziell solche Datenzentren sind daher auch als speichernde und hostende Institutionen eines publikationsbezogenen Datenarchivs in hohem Maße geeignet.

Im Zuge unserer Untersuchungen der Data Policies von wirtschaftswissenschaftlichen Journals evaluierten wir auch die eingesetzte technische Infrastruktur der Datenarchive von Fachzeitschriften. Dabei zeigte sich, dass eine Bereitstellung der Forschungsdaten zu fast 70% über die Webseiten der Zeitschriften erfolgt. Im Regelfall werden die Daten in einem zip-Container als „Supplementary Material“ zur Verfügung gestellt.

---

<sup>15</sup> Aktuell (Anfang August 2012) werden auf den Seiten des RatSWD 17 Forschungsdatenzentren und drei Servicedatenzentren nachgewiesen. <http://ratswd.de/dat/fdz.php>

<sup>16</sup> <http://www.ddialliance.org/>



Einige Zeitschriften nutzen zur Bereitstellung der Daten auch spezielle Webseiten („Data Archives“), auf der die Datensätze aller Ausgaben und Artikel einer Zeitschrift zum Download angeboten werden. Mit Ausnahme einer Übersicht, zu welchen Artikeln Forschungsdaten

existieren, bieten solche speziellen Webseiten aber keine weiteren Features.

Diese Praxis, Forschungsdaten in Form von zip-files in den Anhängen einer Publikation bereit zu stellen, kann als unzureichend angesehen werden, da die Daten nicht zitierbar und nur schwer auffindbar sind. Zudem existieren nur selten Metadaten zu den Forschungsdaten (und wenn, werden dafür keine standardisierten Metadatenschemata verwendet), was eine Nachnutzung der Daten und eine Replikation der Ergebnisse deutlich erschwert. Insbesondere wirkt sich hier negativ aus, dass es keine quodifizierten und von der Community akzeptierten Standards für Replikationen gibt, so dass die Qualität der Beschreibungen zu den Datensätzen teils erheblich variiert.

Etwas anders ist dies bei gut 17% der Fachzeitschriften mit Forschungsdatenrichtlinie, die für ihre Datenarchive spezielle Applikationen wie z.B. Dataverse<sup>17</sup> nutzen, die Funktionalitäten wie persistente Verlinkung und Zitierbarkeit der Forschungsdaten bieten. Über eine Anbindung an Forschungsdatenzentren verfügt hingegen keines der untersuchten Journals.

Um die Optionen von Infrastrukturlösungen für Journal Data Archives weiter zu untersuchen, werden im Rahmen des EDaWaX-Projekt derzeit fast 50 potenzielle Datenprovider dahingehend geprüft, ob diese ein publikationsbezogenes Forschungsdatenarchiv für die Wirtschaftswissenschaften hosten können, wie es gegenwärtig vom EDaWaX-Projekt entwickelt wird.

<sup>17</sup> Die Webseite von Dataverse ist verfügbar unter: <http://www.thedata.org>. Andere Softwarelösungen die ähnliche Funktionalitäten enthalten sind Nesstar (<http://www.nesstar.com>) oder TheDataHub (CKAN) (<http://www.thedatahub.org>).

Neben den FDZ und DSZ, den internationalen Datenarchiven wie ICPSR<sup>18</sup> und CESSDA-Verbund<sup>19</sup> wurden auch einzelne Bibliotheken, Bibliotheksverbände und Archive in die Untersuchung mit einbezogen.

Im Rahmen einer Desktop Research wurde untersucht, welche dieser Institutionen auf die Anforderungen von EDaWaX passen. Speziell wurde geprüft, ob und gegebenenfalls in welcher Form diese Einrichtungen Speichermöglichkeiten für externe Forschungsdatensätze bieten, die für spezifische Publikationen tatsächlich genutzt wurden. Analysiert wurde ferner, welche Richtlinien für die Abgabe von externen Daten gelten, welche technische Infrastruktur dazu eingesetzt wird, welche Metadatenschemata Verwendung finden, ob und wenn wie Nutzer bei der Metadatengenerierung unterstützt werden. Auch die Frage nach ggf. anfallenden Gebühren für Hosting und Speicherung wurde in der Untersuchung berücksichtigt.

Die vorläufigen Ergebnisse der Desktop Recherche zeigten, dass von den untersuchten Institutionen nur eine Minderheit externe Daten annimmt. Zwar gaben 46% der untersuchten Datenzentren an, dass sie generell externe Datensätze annehmen, jedoch beschränkte sich dies oftmals nur auf Datensätze der eigenen Fachdisziplin. Mehr als 10% der untersuchten Datenzentren boten zudem ein Nutzerinterface an, mit dem Wissenschaftler Datensätze über NESSTAR einpflegen können.

Mit der Desktop Research sind die Arbeiten zu diesem Arbeitspaket jedoch noch nicht abgeschlossen. Im Herbst 2012 wird eine Onlinebefragung die Services der Datenzentren detailliert evaluieren.

### **Fazit:**

Im Zuge unserer Studie konnten wir feststellen, dass sich die Zahl der Zeitschriften mit Data Availability Policies in den Wirtschaftswissenschaften deutlich erhöht hat, wenngleich die Anzahl von 29 vor dem Hintergrund von tausenden solcher Zeitschriften eher dem viel zitierten „Tropfen auf den heißen Stein“ gleichkommt, als einem Paradigmenwechsel bei wirtschaftswissenschaftlichen Fachzeitschriften.

Die aufgefundenen Data Policies divergieren erheblich – sowohl hinsichtlich ihres Umfangs als auch hinsichtlich ihrer Qualität. So wird von vielen Richtlinien

---

<sup>18</sup> Interuniversity Consortium for Political and Social Research (ICPSR):  
<http://www.icpsr.umich.edu/icpsrweb/ICPSR/>

<sup>19</sup> Council of European Social Science Data Archives (CESSDA): <http://www.cessda.org/>

beispielsweise der für Replikationen wichtige Berechnungscode nicht mit eingefordert. Andere verzichten auf die Übermittlung von genutzten Programmen, oder verlangen keine dezidierten Beschreibungen der übermittelten Daten und ihrer Funktion. Vor diesem Hintergrund ist anzunehmen, dass ungefähr die Hälfte der untersuchten Richtlinien Replikationen nicht in vollem Umfang gewährleisten dürften. Auf der anderen Seite ist mit der Data Availability Policy der American Economic Review (AER) ein guter Standard geschaffen worden, durch den Replikationen zu einem hohen Prozentsatz erfolgreich verlaufen (Glandon, 2010). Damit, dass 34,5% der untersuchten Richtlinien diese Policy nachnutzen, ist ein Schritt in Richtung einer Best Practise unternommen worden.

Unsere Analysen zeigen aber auch, dass speziell die eingesetzte Infrastruktur zur Bereitstellung von publikationsbezogenen Forschungsdaten bei wirtschaftswissenschaftlichen Fachzeitschriften stark verbesserungswürdig ist. Gegenwärtig werden Forschungsdaten zumeist als zip-Datei an die Publikationen angehängt und sind über die Webseiten des Journals verfügbar. Da zumeist keine persistenten Identifikatoren vergeben werden und keine weiteren Metadaten zu den Daten hinzugefügt werden, sind die Forschungsdaten oft nicht zitier- und suchbar.

Zudem erfolgen auf den Servern der Herausgeber oder der Verlage keine Maßnahmen zum Datenmanagement oder zum Langzeiterhalt der Daten.

Eine wesentlich bessere Praxis wird von einer Minderheit von Journals betrieben, die Datenrepositories wie Dataverse für ihre Datenarchive nutzen. Hier finden sich sowohl persistente Identifikatoren als auch ein rudimentäres Set an Metadaten. Dennoch bieten auch diese publikationsbezogenen Forschungsdatenarchive keine Anbindung an größere Forschungsdatenzentren, die eine professionelle Datenvorhaltung gewährleisten könnten.

Im Rahmen des EDaWaX-Projekts werden wir in den nächsten Monaten hier ansetzen und das Potential einer Kooperation zwischen Fachzeitschriften und Forschungsdatenzentren im Rahmen von Umfragen und Interviews näher beleuchten.

#### **Literaturverzeichnis**

- Allianz der deutschen Wissenschaftsorganisationen: Grundsätze zum Umgang mit Forschungsdaten, Berlin 2010. Verfügbar unter: [http://www.ratswd.de/download/RatSWD\\_WP\\_2010/RatSWD\\_WP\\_156.pdf](http://www.ratswd.de/download/RatSWD_WP_2010/RatSWD_WP_156.pdf)
- Anderson, R./Greene, W./McCullough, B.D./Vinod, H. D.: The Role of Data/Code Archives in the Future of Economic Research. In: Journal of Economic Methodology, Nr.1, S. 99-119.



- Bräuninger, Michael/Haucap, Justus/Muck, Johannes: Was lesen und schätzen Ökonomen im Jahr 2011?, DICE Ordnungspolitische Perspektiven, Nr. 18. Verfügbar unter: <https://econstor.eu/dspace/bitstream/10419/49023/1/667448497.pdf>
- Dewald, William G./Thursby, Jerry G./Anderson, Richard G.: Replication in Empirical Economics: The Journal of Money, Credit and Banking Project. In: The American Economic Review, Vol. 76, No. 4 (1986), S. 587-603.
- Glandon, Phillip: Report on the American Economic Review Data Availability – Compliance Project, Vanderbilt University, November 2010. Verfügbar unter: [http://www.aeaweb.org/aer/-2011\\_Data\\_Compliance\\_Report.pdf](http://www.aeaweb.org/aer/-2011_Data_Compliance_Report.pdf)
- Huschka, Denis/Oellers, Claudia/Ott, Notburga/Wagner, Gert G.: Datenmanagement und Data Sharing: – Erfahrungen in den Sozial- und Wirtschaftswissenschaften. In: Büttner, Stephan / Hobohm, Hans-Christoph /Müller, Lars: Handbuch Forschungsdatenmanagement, Bad Honnef, 2011, S. 35 - 48. Verfügbar unter: <http://opus4.kobv.de/opus4-fhpotsdam/files/208/HandbuchForschungsdatenmanagement.pdf>
- King, Gary: Replication, Replication. In: PS: Political Science and Politics, Nr. 28, 1995, S. 443-499. Verfügbar unter: <http://gking.harvard.edu/gking/files/replication.pdf>
- Mauer, Reiner: Das GESIS Datenarchiv für Sozialwissenschaften. Vortrag beim Workshop „Archivierung sozial- und wirtschaftswissenschaftlicher Datenbestände“ in der Deutschen Nationalbibliothek, Frankfurt, 15./16. September 2011. Verfügbar unter: [http://www.ratswd.de/ver/-docs\\_Archivierung\\_2011/mauer.pdf](http://www.ratswd.de/ver/-docs_Archivierung_2011/mauer.pdf)
- McCullough, B.D.: Got Replicability? The Journal of Money, Credit and Banking Archive. In: Econ Journal Watch, Nr. 4, 2007, S. 326-337.
- McCullough, B.D.: Open Access Economics Journals and the Market for Reproducible Economic Research. In: Economic Analysis and Policy, Nr. 39, 2009, S. 117-126.
- McCullough, B.D./McGeary, K. A. / Harrison, T.: Lessons from the JMCB Archive. In: Journal of Money, Credit, and Banking, Nr. 38, 2006, S. 1093-1107.
- McCullough, B.D./McGeary, Kerry Anne / Harrison, T.: Do economics journal archives promote replicable research? In: Canadian Journal of Economics, Nr. 41, 2008, S. 1406-1420.
- McCullough, B.D./McKittrick, Ross: Check the Numbers: The Case for Due Diligence in Policy Formation, Studies in Risk and Regulation, Fraser Institute, 2009. Verfügbar unter: <http://www.fraserinstitute.org/WorkArea/DownloadAsset.aspx?id=2486>
- McCullough, B.D./Vinod, H.D.: Verifying the Solution from a Nonlinear Solver: A Case Study." In: American Economic Review, Vol. 93, Issue 3, S. 873-892. DOI:10.1257/000282803322157133
- Reilly, S. / Schallier, W. / Schrimpf, S. / Smit, E. / Wilkinson, M.: Opportunities for Data Exchange: Report on Integration of Data and Publications, Oktober 2011. Verfügbar unter: <http://www.alliancepermanentaccess.org/wp-content/plugins/download-monitor/download.php?id=ODE+Report+on+Integration+of+Data+and+Publications>
- Wagner, Gert G. / Huschka, Denis: Datenverfügbarkeit reicht nicht, um Replikationsstudien zur Routine zu machen. RatSWD Working Paper Series Nr. 194, 2012. Verfügbar unter: - [http://www.ratswd.de/download/RatSWD\\_WP\\_2012/RatSWD\\_WP\\_194.pdf](http://www.ratswd.de/download/RatSWD_WP_2012/RatSWD_WP_194.pdf)
- Wissenschaftsrat: Empfehlungen zu Forschungsinfrastrukturen, Köln, 2011. Verfügbar unter: <http://www.wissenschaftsrat.de/download/archiv/10464-11.pdf>
- Wissenschaftsrat : Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020, Berlin, 2012. Verfügbar unter: <http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf>