

Forschungsdatenpolicies in wissenschaftlichen Zeitschriften – Eine empirische Untersuchung

Foto: Laura Waldherr



Benedikt Blahous

Foto: Frank Helmrich



Juan Gorraiz

Foto: Frank Helmrich



Christian Gumpenberger

Foto: Julian Bernhardt



Oliver Lehner

In den Naturwissenschaften, aber auch in Fachgebieten der Sozialwissenschaften, die vorwiegend empirische Methoden verwenden, werden neue Ergebnisse überwiegend in Zeitschriften veröffentlicht. Dabei spielen die Veröffentlichungsrichtlinien für Autoren eine zentrale Rolle, welche mittlerweile fakultativ auch Policies für das Hinterlegen der Forschungsdaten enthalten können, um die publizierten Ergebnisse zu stützen. In dem hier vorgestellten Projekt »Richtlinien wissenschaftlicher Zeitschriften für das Hinterlegen von Forschungsdaten« ging es um das Identifizieren von Zeitschriften mit einer Data Policy mittels verschiedener Suchstrategien und der Bewertung derselben hinsichtlich Effizienz sowie der Einstufung der vorgefundenen Data Policies bezüglich ihrer Stärke. Von 534 untersuchten Zeitschriften quer durch alle Disziplinen hatten insgesamt 346 eine Data Policy mit folgenden Abstufungen: 68 (12,7%) strong, 67 (12,5%) weak und 211 (39,5%) optional.

In the sciences as well as in some social sciences disciplines which mainly rely on empirical methodologies, journals are the preferred publication channel. In this respect author guidelines are of major importance. These may also include deposit policies concerning supportive research data. The »Richtlinien wissenschaftlicher Zeitschriften für das Hinterlegen von Forschungsdaten« project deals with the identification of journals with available data policies using various search strategies. These strategies are evaluated with regard to their efficiency, and the identified data policies are categorized according to their strength. Of the 534 cross-disciplinary journals analysed, 346 were identified with an available data policy which could be categorized as follows: 68 (12.7%) strong, 67 (12.5%) weak and 211 (39.5%) optional.

EINLEITUNG

In den letzten Jahren hat sich im Wissenschaftsbetrieb das Bewusstsein für die Wichtigkeit der Langzeitarchivierung von Forschungsdaten verstärkt. Dies belegen zahlreiche Initiativen und Projekte, die zu diesem Thema ins Leben gerufen wurden und bei denen auch wissenschaftliche Bibliotheken beteiligt sind. Aktuelle Beispiele dafür sind das mit Anfang 2014 gestartete Projekt *e-Infrastructures Austria*, außerdem die internationale Initiative *Research Data Alliance*, das EU-Projekt *Horizon 2020* und das Förderprogramm *Informationsstrukturen für Forschungsdaten* der Deutschen Forschungsgemeinschaft.¹ Die Langzeitarchivierung von Forschungsdaten ist deshalb von großer Bedeutung, weil dadurch:

- die Aussagen in wissenschaftlichen Publikationen, die sich auf diese Daten beziehen, langfristig überprüfbar sind,
- Untersuchungen mit großen Datenmengen erleichtert werden (z. B. Langzeitstudien),
- Daten in der Forschung wiederverwendet werden können – sofern sie für andere Forscher² zugänglich sind (vgl. Oßwald et al., 2012, S. 13 f.).

Der Terminus »Forschungsdaten« ist inhaltlich schwer zu bestimmen, bisher gibt es keine einheitliche Definition. Dies liegt u. a. an den vielfältigen Forschungsgegenständen und -methoden in den verschiedenen Fachgebieten, und an dem Umstand, dass prinzipiell alles ein Forschungsgegenstand werden kann (vgl. Ludwig, 2012, S. 300). Erläuterungsversuche beziehen sich häufig darauf, welche Rolle die Daten in der Forschung spielen. Sánchez Solís charakterisiert den entsprechenden englischen Ausdruck »research data« z. B. folgendermaßen: »Some people refer to data as anything digital which is the product of research, whereas others refer to data as any material that is collected, observed or created either to produce original research results or to validate original research results. In this respect, the term »research data« means data on which an argument or hypothesis is based. Data may be raw or processed and may be stored and shared in any format or media. Digital research data can be regarded as data created in a digital form (born digital) or converted to a digital form (digitised).« (Sánchez Solís, 2014, S. 4f.) Forschungsdaten sind also Daten, die einer wissenschaftlichen Aussage oder Hypothese zugrunde liegen: z. B. chemische Strukturen, Chromatogramme, Daten aus neurologischen bildgebenden Verfahren, Gensequenzen, ökonomische oder klimatische Modelle, Sensordaten und Umfrageergebnisse (vgl. Sánchez Solís, 2014, S. 6).

In den Naturwissenschaften, aber auch in Fachgebieten der Sozialwissenschaften, die vorwiegend empirische Methoden verwenden, werden neue Ergebnisse überwiegend in Zeitschriften veröffentlicht. Deshalb haben Zeitschriften zum Teil erheblichen Einfluss darauf, wie Forscher ihre Ergebnisse publizieren. Dabei spielen die Veröffentlichungsrichtlinien für Autoren eine zentrale Rolle. Diese betreffen i. d. R. den Inhalt, den Aufbau, die Formatierung und den Ablauf des Einreichungsprozesses eines Aufsatzes. Um seine Ergebnisse in einer Zeitschrift veröffentlichen zu können, muss ein Forscher diese Richtlinien einhalten. Ein Teil der Zeitschriften formuliert inzwischen außerdem Richtlinien oder Empfehlungen für das Hinterlegen von zusätzlichen Daten, insbesondere für das Hinterlegen der Forschungsdaten, welche die publizierten Ergebnisse stützen. Da die Zeitschriften durch Data Policies den Umgang der Forscher mit ihren Forschungs-

daten direkt beeinflussen, ist ein Überblick über diese von großem Interesse. Dazu gibt es bisher aber nur wenige Informationsquellen.

In dem hier vorgestellten Projekt »Richtlinien wissenschaftlicher Zeitschriften für das Hinterlegen von Forschungsdaten« an der Universität Wien³ ging es nun darum, Zeitschriften mit einer Data Policy zu identifizieren und diese Data Policies zu analysieren, vor allem im Hinblick auf die (Langzeit)Archivierung. Die gesammelten Informationen sollten einen Eindruck vom Status quo vermitteln und als Grundlage für weitere Untersuchungen dienen können. Die Teilziele dieser Studie waren:

- die Bestandsaufnahme von verwandten Projekten (insbesondere JoRD),
- die Formulierung und Bewertung von Suchstrategien zum Finden von Zeitschriften mit einer Data Policy,
- das Identifizieren einer repräsentativen Anzahl (> 200) von Zeitschriften mit einer Data Policy sowie die Analyse der Data Policies,
- die Dokumentation und Diskussion der Ergebnisse.

BESTANDSAUFNAHME

Am Beginn dieser Studie stand die Analyse von *Journal Research Data Policy Bank* (JoRD)⁴, einem Projekt der University of Nottingham, welches vom *Joint Information Systems Committee* finanziert wurde. JoRD war eine Machbarkeitsstudie zu der Frage, wie ein Informationsservice (eine Datenbank) beschaffen sein sollte, das Data Policies sammelt und strukturiert darstellt. Die angepeilten Zielgruppen waren Forscher sowie Verlage, Geldgeber und Betreiber von Repositorien.

Im Zeitraum von sechs Monaten (Juli bis Dezember 2012) wurden ca. 400 wissenschaftliche Zeitschriften analysiert und vorhandene Data Policies gesammelt und bewertet. Außerdem wurden Umfragen und Interviews mit Vertretern aller Interessengruppen durchgeführt. Im Rahmen von JoRD wurden jeweils die 100 ersten und 100 letzten Zeitschriften der Thomson Reuters *Journal Citation Reports* 2012 (geordnet nach Impact Factor) für die Bereiche Sciences und Social Sciences analysiert. Der Bereich Sciences umfasst u. a. die Fachgebiete Biology, Chemistry, Engineering, Materials Science und Physics. Social Sciences umfasst u. a. die Fachgebiete Geography, Criminology, Economy, Psychology und Law.

Data Policies wurden als »strong« kategorisiert, wenn das Hinterlegen von Daten als notwendige Voraussetzung für die Veröffentlichung eines wissenschaftlichen Aufsatzes genannt wurde und als »weak«, wenn die Autoren zum Hinterlegen von Daten nur aufgefordert wurden. Die Analyse der Zeitschriften ergab,

dass strong Data Policies eher bei Zeitschriften mit einem hohen Impact Factor, weak Data Policies eher bei Zeitschriften mit einem niedrigen Impact Factor zu finden sind. Der Vergleich zwischen Sciences und Social Sciences zeigt:

- In beiden Bereichen gibt es überwiegend weak Data Policies.
- Im Bereich Sciences gibt es deutlich mehr Zeitschriften mit einer Data Policy als im Bereich Social Sciences.
- Der Anteil an strong Data Policies ist im Bereich Sciences etwas höher. Zeitschriften aus diesem Bereich enthalten auch eher Verweise zu Fachrepositorien, wo Daten wie Massenspektren, Nukleotid- und Proteinsequenzen und chemische Strukturen hinterlegt werden (vgl. Bamkin et al., 2012, 13.09.2012 und 15.10.2012).

Die Umfragen und Interviews ergaben, dass bei allen Interessengruppen (Forscher, Verlage, Geldgeber, Betreiber von Repositorien) die Einstellung zum Hinterlegen von Forschungsdaten grundsätzlich positiv ist, was sich bis dato allerdings noch nicht in der Praxis zeigt. Aufgrund der Umfrageergebnisse wurden von JoRD zwei Hindernisse identifiziert (Bamkin et al., 2012, 04.01.2014): »First the stereotype of the data hugging scientist must disappear from the minds of researchers, funders, Higher Educational Institutions and publishing houses. Secondly, the infra-structure of data deposit sites, how, when and where to deposit data, has to be fully resolved, publicised and implemented.«

Forscher meinen, dass Projekte, die fremde Daten verwenden, weniger Geldmittel von Universitäten und Drittmittel von anderen Organisationen erhielten und neue Projekte bevorzugt würden. Neben praktischen Gründen wie Zeitmangel und Mangel an Ressourcen wüssten Forscher aufgrund von fehlenden Standards oft nicht, wo und wie sie Daten hinterlegen sollen. Forscher geben an, dass sie den Aufbau von Informationsservices zu Data Policies, wie von JoRD angedacht, begrüßen würden.

Die Analyse von JoRD wurde in erster Linie anhand des Projekt-Blogs durchgeführt, in welchem der Ablauf umfassend dokumentiert ist. Zusätzlich wurde Kontakt zum Team von JoRD aufgenommen, um offene Fragen zu klären und Hintergrundinformationen zum Projekt zu erhalten. Das Team war sehr kooperativ und stellte bereitwillig die vollständige Liste der analysierten Zeitschriften, Statistiken, Grafiken und Literaturhinweise zur Verfügung. Die Auseinandersetzung mit JoRD ermöglichte einen guten Einstieg in die Thematik, und die Kooperation mit dem englischen Projektteam war bei der Planung der vorliegenden Projektarbeit von großer Hilfe. Die Vorgehensweise von JoRD



Bettina Stein



Ursula Ulrych

Journal Research Data Policy Bank (JoRD)

Analyse der Zeitschriften

erwies sich als wertvolle Orientierung für die Recherche von relevanten Zeitschriften und für die Analyse der identifizierten Data Policies. Die JoRD-Analyse ist somit als Voruntersuchung für die eigentlich durchgeführte Studie zu verstehen, die nachfolgend beschrieben wird.

METHODIK

Suchstrategien

Aus Mangel an Verzeichnissen mit den entsprechend relevanten Daten wurden nach der Bestandsaufnahme verschiedene Suchstrategien zum Identifizieren von Zeitschriften mit einer Data Policy formuliert und getestet. Die identifizierten Zeitschriften sollten einen Eindruck davon vermitteln, welche Arten von Data Policies es gegenwärtig in (englischsprachigen) Zeitschriften gibt und wie stark diese jeweils vertreten sind. Es wurde darauf geachtet, eine breit gefächerte Auswahl an Fachgebieten abzudecken. Berücksichtigt wurden dabei besonders auch geistes- und sozialwissenschaftliche Fachgebiete. Außerdem sollte in jedem Fachgebiet in etwa dieselbe Anzahl an Zeitschriften analysiert werden, mindestens jedoch 20 Zeitschriften. Eine optimale Suchstrategie sollte als Ergebnis eine dublettenfreie Liste von Zeitschriften mit einer Data Policy und einem Link zur Website der Zeitschrift haben. Bei der Formulierung der Suchstrategien wurde darauf geachtet, ein möglichst breites Spektrum von Datenquellen zu berücksichtigen. Im Folgenden sind diese Quellen mit der entsprechenden Vorgehensweise überblicksmäßig angeführt. Die Datenerhebung wurde vom 09.05.2014 bis zum 27.05.2014 durchgeführt.

Google Scholar

Vorüberlegung: Google Scholar indiziert eine sehr große Menge an wissenschaftlicher Literatur und ist einfach zu verwenden. Deshalb bietet sich diese Datenquelle für eine erste Suche an.

Vorgehensweise: Volltext- und Titelsuchen mit den Begriffen ›data‹, ›policy‹, ›scientific journal‹, ›research data‹ und verschiedenen Kombinationen davon wurden durchgeführt. Es wurden nur solche Suchergebnisse weiterverfolgt, die anhand der Vorschau als möglicherweise relevant eingestuft wurden.

Repositorien

Vorüberlegung: Zeitschriften, die einen Aufsatz beinhalten, zu dem bereits Daten in einem Repositorium hinterlegt wurden, haben vermutlich eine Data Policy. *Vorgehensweise:* Die Quelle für Repositorien war die *Registry of Research Data Repositories*, die Metadaten

zu mehr als 690 Repositorien enthält. Die Repositorien wurden darauf geprüft, ob sie Verweise auf Zeitschriften enthalten: War eine Liste von Zeitschriften, in denen auf das Repositorium verwiesen wird, verfügbar, wurde diese systematisch abgearbeitet. Fehlte eine entsprechende Liste, wurde im Repositorium eine erste Suche mit einem Schlagwort aus dem Fachgebiet durchgeführt und die Metadaten der gefundenen Datensätze auf Referenzen zu Zeitschriften geprüft. Enthielten die Metadaten entsprechende Verweise, so wurden mit einer weiteren Suche die Treffermenge möglichst vergrößert, die Verweise aus den Metadaten gesammelt und abgearbeitet.

Zeitschriften-Rankings

Vorüberlegung: Zeitschriften mit einem hohen Impact Factor haben mit großer Wahrscheinlichkeit eine Data Policy (vgl. Piwowar et al., 2008, S. 2). Das gilt vermutlich auch für alle mit den *Journal Citation Reports* vergleichbaren Rankings, welche in erster Linie darauf basieren, dass gezählt wird, wie oft die Artikel, die in einer Zeitschrift veröffentlicht wurden, in einem bestimmten Zeitraum zitiert wurden.

Vorgehensweise: Als Quelle diente *SCImago Journal & Country Rank*, der die Erstellung von Rankings für einzelne Fachgebiete erlaubt. Nach Auswahl eines Fachgebietes wurden jeweils die ersten 20 Zeitschriften analysiert, sofern zumindest fünf der zehn top-gearankten Zeitschriften eine Data Policy hatten. In einigen Fällen wurden auch mehr als 20 Zeitschriften analysiert.

Websites von Fachverlagen

Vorüberlegung: Findet man eine Zeitschrift mit einer Data Policy, so gibt es mit hoher Wahrscheinlichkeit weitere Zeitschriften desselben Verlages, die auch eine Data Policy haben. Selbstverständlich ist dies bei den Zeitschriften letztlich von den Herausgebern abhängig und muss somit nicht zwangsläufig für alle Zeitschriften eines Verlages gelten.

Vorgehensweise: Hat eine vorliegende Zeitschrift eine strong oder weak Data Policy, wurde geprüft, ob es sich um eine Zeitschrift eines Fachverlages mit einer überschaubaren Anzahl an Zeitschriften handelt. In diesem Fall wurden auch die restlichen Zeitschriften des Verlages analysiert.

Auf eine zusätzliche Autopsie von physischen Zeitschriftenheften bei (auch) im Printformat vorliegenden Zeitschriften wurde bewusst verzichtet. Einerseits hätte dieser Aufwand den zeitlichen Rahmen dieser Studie gesprengt, andererseits ist davon auszugehen, dass vorhandene Richtlinien auch elektronisch publiziert sein sollten.

geistes- und sozialwissenschaftliche Fachgebiete

Spektrum der Datenquellen

Datenerfassung

Bei allen Zeitschriften wurden die folgenden Daten erfasst:

- die Suchstrategie, mit der die Zeitschrift gefunden wurde,
- das Fachgebiet der Zeitschrift,
- der Titel, die ISSN der Druckausgabe, die ISSN der Online-Ausgabe und der Verlag,
- die URL der Website der Zeitschrift,
- das Bearbeitungsdatum,
- ob die Zeitschrift einen Eintrag im *Directory of Open Access Journals* (DOAJ) hatte und ob eine Data Policy vorhanden war.

Bei den Zeitschriften mit einer Data Policy wurde zusätzlich die Stärke der Data Policy bestimmt und der Link zur Data Policy, die Kontaktdaten der Redaktion der Zeitschrift und gegebenenfalls auch in der Data Policy explizit angeführte Datentypen, Datenformate und Repositorien erfasst. Die Informationen zu Titel, ISSN und Verlag stammen jeweils entweder aus *Ulrichsweb*, einem internationalen Zeitschriftenverzeichnis mit detaillierten bibliografischen Angaben, oder direkt von der Website der Zeitschrift. War der in *Ulrichsweb* angegebene Link veraltet, so wurde die Website der Zeitschrift in Google gesucht (Suche nach Titel oder ISSN). Die Kontaktdaten wurden ausnahmslos über die Websites der Zeitschriften ermittelt. Die Indexierung im DOAJ wurde für jede Zeitschrift mittels Suche der Online-ISSN direkt im DOAJ geprüft. Datenformate und Repositorien wurden bei entsprechendem Vorhandensein einer Data Policy derselben direkt entnommen. Im Folgenden werden die oben genannten Attribute, welche die Basis für die durchgeführten Analysen zur Stärke der Data Policies bildeten, erläutert.

Fachgebiet

Die Basis der Zuordnung der Zeitschriften zu Fachgebieten bildeten die Angaben in *Ulrichsweb*, die Bezeichnungen der Fachgebiete in *SCImago Journal & Country Rank* sowie die auf den Websites der Zeitschriften beschriebenen ›Aims and Scopes‹. Dabei ergaben sich mehrere Schwierigkeiten:

- Das zugeordnete Fachgebiet in diesen Quellen war oft zu spezialisiert. In solchen Fällen wurde ein übergeordnetes Fachgebiet vergeben, z. B. *Chemistry* statt *Organic Chemistry*.
- Falls eine Zeitschrift multidisziplinär oder mehreren Fachgebieten zugeordnet war, wurde intellektuell das vermutlich vorrangige Fachgebiet ausgewählt.
- Die Fachgebiete *Agriculture*, *Biotechnology* und

Food Science wurden zu *Applied Sciences* zusammengefasst.

Für die Ermittlung der prozentualen Verteilung der verschiedenen Stärken der Data Policies in den Fachgebieten wurde die Anzahl der Zeitschriften pro Fachgebiet auf 27 – das ist der Median der absolut pro Fachgebiet analysierten Zeitschriften – normiert und die Häufigkeiten der Data Policies entsprechend skaliert, um einen direkten Vergleich der Fachgebiete zu ermöglichen.

Verlag

Der Verlagsname wurde aus *Ulrichsweb* oder direkt von der Website der Zeitschrift übernommen. Für eine bessere Vergleichbarkeit wurden die Verlagsnamen normiert, insbesondere Angaben zum Verlagsort und bezüglich der Gesellschaftsform wurden weggelassen, z. B. wurde ›*Springer New York LLC*‹ zu ›*Springer*‹ oder ›*Elsevier BV*‹ zu ›*Elsevier*‹. Dagegen wurden bei Zeitschriften die Verlagsnamen unverändert übernommen, die weiterhin unter der Marke früher eigenständiger Verlage veröffentlicht werden, z. B. ›*Pergamon Press*‹ statt ›*Elsevier*‹.

Data Policy

Data Policies fanden sich – wenn vorhanden – meist als Abschnitt in den Autorenrichtlinien. Seltener lagen sie eigenständig vor. In jedem Fall war das entsprechende HTML- oder PDF-Dokument ausgehend von der Website der Zeitschrift zu finden. Die Identifikation des gesuchten Dokuments war im Allgemeinen problemlos, auch wenn die Bezeichnungen nicht einheitlich waren: ›*Guide for Authors*‹, ›*Instructions*‹, ›*Manuscript Submission Guidelines*‹ und Ähnliches.

Selbstverständlich ist es möglich, dass bei den analysierten Zeitschriften existierende Richtlinien bezüglich des Umgangs mit Forschungsdaten nicht auf den Websites publiziert sind und nur durch persönlichen Kontakt zu den Redaktionen oder Herausgebern in Erfahrung gebracht werden könnten. Eine solche Erhebung hätte aber einerseits den Rahmen dieser Studie gesprengt, andererseits können nur explizit öffentlich publizierte Richtlinien auch als wirkliche Policies angesehen werden.

Stärke der Data Policy

Jede identifizierte Data Policy wurde hinsichtlich ihrer Stärke bewertet und einer von drei Kategorien zugeordnet: *strong*, *weak* oder *optional*. Das Kategorisierungsmerkmal war, wie deutlich das Hinterlegen von Daten in der Data Policy gefordert wurde (Verpflichtung, Aufforderung oder Möglichkeit). Ursprünglich waren in Anlehnung an JoRD nur zwei Kategorien

bibliografische Angaben

Kategorisierungsmerkmal

Ecological Monographs

The pages of *Ecological Monographs* are open to integrative original research papers and synthetic reviews. As a condition for publication of a manuscript in *Ecological Monographs*, all data associated with the results reported in manuscripts must be made available in a permanent, publicly accessible data archive or repository. Policy for manuscripts submitted on or after January 1, 2011. [more]

Data Policy. As a condition for publication of a manuscript in *Ecological Monographs*, all data associated with the results reported in manuscripts must be made available in a permanent, publicly accessible data archive or repository. Examples of permanent data repositories include GenBank for DNA sequences (<http://www.ncbi.nlm.nih.gov/genbank/>), *Ecological Archives* (<http://esapubs.org/archives/>) for basic ASCII tables, and ORNL-DMAC (<http://daac.ornl.gov/>) for biogeochemical data. *Ecological Monographs* also is a partner with Dryad (<http://datadryad.org/>), which provides a flexible platform for a wide variety of digital data. With the exception of material submitted to *Ecological Archives*, data do not have to be archived at the time of manuscript submission, only following acceptance of the manuscript for publication.

Abb. 1: Beispiel für eine strong Data Policy

SUPPLEMENTARY MATERIAL

Authors are encouraged to submit supplementary materials that are directly related to their papers. The supplements are subject to editorial review and approval. More information on supplements is available at <http://journals.taylorandfrancis.com/amstat/supplementary-materials/>.

SUBMISSION OF DATA SOURCES AND CODE

The ASA strongly encourages authors to submit datasets, code, other programs, and/or extended appendices that are directly relevant to their submitted articles. These materials are valuable to users of the ASA's journals and further the profession's commitment to reproducible research.

Whenever a dataset is used, its source should be fully documented and the data should be made available as an online supplement. Exceptions for reasons of security or confidentiality may be granted by the Editor. Whenever specific code has been used to implement or illustrate the results of a paper, that code should be made available if possible.

Abb. 2: Beispiel für eine weak Data Policy

Beispiele für Data Policies

(strong und weak) geplant gewesen. Bei der Analyse von Data Policies in der Testphase der Suchstrategien stellte sich aber heraus, dass viele Zeitschriften Empfehlungen für das Hinterlegen von Daten hatten, die zu »schwach« waren, um als weak Data Policies gelten zu können. In diesen Empfehlungen wurden die Autoren nicht explizit aufgefordert, Daten zu hinterlegen, sondern nur auf die Möglichkeit hingewiesen, dies zu tun. Außerdem konnten auch Daten hinterlegt werden, die keinen direkten Bezug zu den Ergebnissen des eingereichten Aufsatzes hatten. Das Vorhandensein dieser sehr schwachen Data Policies war aber eine wichtige Information, die ohne eine dritte Kategorie verloren gegangen wäre. Deshalb wurde zusätzlich die Kategorie optional eingeführt.⁵ Im Folgenden werden

die drei Kategorien strong, weak und optional anhand von Beispielen erläutert.

Strong Data Policy

Eine strong Data Policy verlangt als Voraussetzung für die Veröffentlichung eines Artikels die Offenlegung der zugrunde liegenden Forschungsdaten. Die Formulierung der Data Policy legt nahe, dass der entsprechende Aufsatz andernfalls nicht veröffentlicht wird. Abbildung 1 zeigt als Beispiel den Screenshot der Data Policy der Ecological Society of America (<http://esapubs.org/esapubs/emonTypes.htm#Dat>, 27.05.2014).

Weak Data Policy

Eine weak Data Policy zeigt eindeutigen Aufforde-

rungscharakter, die Forschungsdaten zu hinterlegen. Im Gegensatz zu einer strong Data Policy ist dies jedoch keine notwendige Voraussetzung, um einen Artikel veröffentlichen zu können. Abbildung 2 zeigt als Beispiel den Screenshot der Data Policy der Zeitschrift *Journal of Business & Economic Statistics* (www.tandfonline.com/action/authorSubmission?journalCode=ubes20&page=instructions, 23.12.2014).

Optional Data Policy

Eine optional Data Policy weist auf die Möglichkeit hin, zusätzliche Daten zu hinterlegen, überlässt aber dem Autor die Entscheidung, ob und welche Daten zur Verfügung gestellt werden. Hierbei kann es sich auch um rein unterstützende Materialien handeln (Audiodateien, Illustrationen, etc.), die normalerweise nur in der Online-Version als »supplemental material« angeboten werden. Abbildung 3 zeigt als Beispiel den Screenshot der Data Policy einer Elsevier Zeitschrift (www.elsevier.com/journals/applied-geography/0143-6228/guide-for-authors#87000, 27.05.2014).

Keywords

Im Hinblick darauf, dass es bislang keine allgemeingültige Definition des Ausdrucks »Forschungsdaten« gibt, wurde außerdem untersucht, wie die zu hinterlegenden Daten in den Data Policies bezeichnet werden und ob dabei auch mit »Forschungsdaten« synonyme Ausdrücke wie »research data«, »research material« und »scientific data« vorkommen. Dazu wurde jede identifizierte Data Policy gelesen und alle Ausdrücke, welche auf die zu hinterlegenden Daten Bezug nahmen, intellektuell ausgewählt. Für die geplante quantitative Analyse wurden diese Ausdrücke normiert im Nominativ Singular gesammelt. Außerdem wurden zwei Sammelbegriffe für längere Textpassagen eingeführt, die inhaltlich relevant, aber zu komplex waren, um in der oben beschriebenen Form normiert zu werden: »specific« und »replication material«. Der Sammelbegriff »specific« wurde vergeben, wenn in einer Data Policy explizit das Hinterlegen bestimmter Datentypen, z. B. DNA-Sequenzen oder chemische Strukturen, gefordert (oder gewünscht) wurde. Der Sammelbegriff »replication material« wurde für Phrasen vergeben, in denen sinngemäß gefordert (oder gewünscht) wurde, dass genau die Daten hinterlegt werden sollten, welche die präsentierten Ergebnisse reproduzierbar und überprüfbar machen. Die Gesamtheit der normierten Ausdrücke und der beiden Sammelbegriffe (zusammen kurz »Keywords«), welche für die untersuchten Data Policies vergeben wurden, bildeten die Grundlage für die quantitative Analyse.

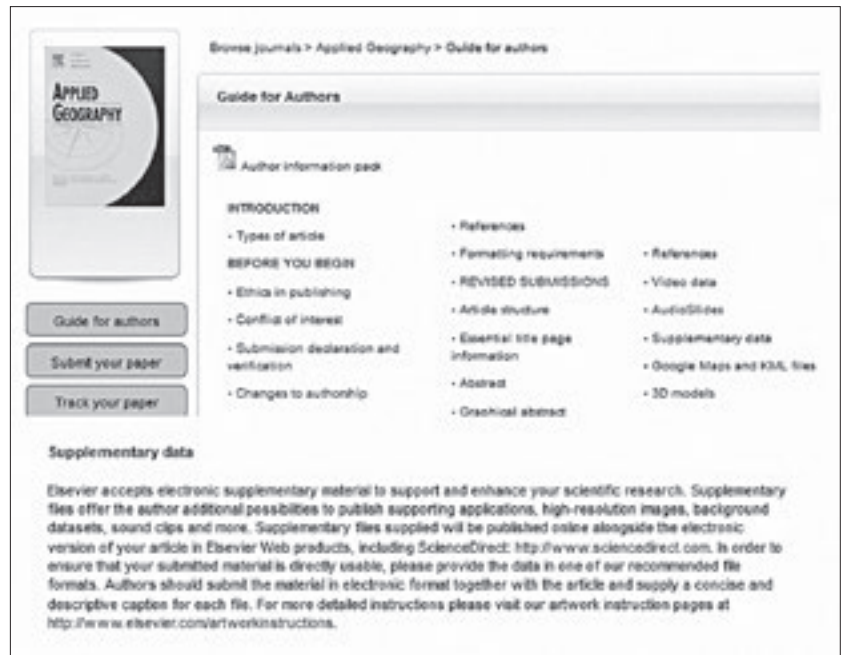


Abb. 3: Beispiel für eine optional Data Policy

ERGEBNISSE

Suchstrategien

Drei der vier Quellen wurden nach einer Testphase für die Suche genutzt: Repositorien, Websites von Fachverlagen und Zeitschriften-Rankings. Es wurden 559 Zeitschriften von 112 Verlagen gefunden, davon waren 25 Dubletten von JoRD. In der Datenanalyse wurden auch diejenigen Zeitschriften berücksichtigt, die während der Testphase gefunden worden waren, die Dubletten von JoRD wurden nicht berücksichtigt. Von den anderen 534 Zeitschriften stammten 271 (50,8%) aus Zeitschriften-Rankings, 128 (24,0%) aus Repositorien und 123 (23,0%) von Websites von Fachverlagen. Die höchste Trefferquote gab es bei Websites von Fachverlagen: hier hatten 99,2% aller gefundenen Zeitschriften eine Data Policy. Die Trefferquote in Repositorien war 61,7%, in Zeitschriften-Rankings 50,2%. Den größten Anteil an weak und strong Data Policies gab es in Repositorien: Hier waren 59,5% der gefundenen Data Policies strong oder weak, gefolgt von Websites von Fachverlagen (38,5%), und Zeitschriften-Rankings (27,9%). Während der Testphase wurden zwölf Zeitschriften in Google Scholar gefunden. Die Trefferquote in Google Scholar war 75,0%, der Anteil an weak und strong Data Policies lag bei 33,3%.

Die Trefferquote in Repositorien hing stark vom Fachgebiet ab. In naturwissenschaftlichen Fachgebieten war diese sehr hoch: Die Trefferquote in *Dryad* war 100% (davon 70% strong Data Policies) und in der *Crystallography Open Database* bei 87,5% (davon 50% strong Data Policies). In Fachgebieten außerhalb der

Keywords-Sammlung

Trefferquoten

Name der Liste	1–10	11–20	1–20
Geology	7	9	16 (80 %)
Law	3	2	5 (25 %)
Linguistics	6	4	10 (50 %)
Music	5	6	11 (55 %)
Pharmaceutical Science	7	9	16 (80 %)
Sociology and Political Science	7	9	16 (80 %)

Tab. 1: Anzahl der Zeitschriften mit einer Data Policy in den angegebenen Bereichen der Listen aus dem *SCImago Journal & Country Rank*



Abb. 4: Anzahl der Zeitschriften nach der Stärke der Data Policies



Abb. 5: Anzahl der Zeitschriften nach Fachgebiet

Naturwissenschaften war die Trefferquote deutlich geringer: Im *International Food Policy Research Institute Dataverse* lag diese bei 51,1%, im *National Archive of Criminal Justice Data* nur bei 28,6%. Die Anzahl der Zeitschriften auf den verwendeten Websites von Fachverlagen variierte stark und reichte von drei (*Society for American Archaeology*) bis 71 (*American Psychological Association*). Innerhalb eines Verlages waren die Data Policies meist einheitlich.

Es wurden elf Listen aus dem *SCImago Journal & Country Rank* 2012 analysiert: Archaeology, Astronomy and Astrophysics, Computer Vision and Pattern Recognition, Earth and Planetary Sciences, Economics, Geology, Law, Linguistics, Music, Pharmaceutical Science und Sociology and Political Science. Es wurden mindestens 20 (Economics, Law, Pharmaceutical Science) und maximal 43 Zeitschriften (Computer Vision and Pattern Recognition) analysiert. Die Trefferquote schwankte zwischen 25 % in Law und 80 % in Pharmaceutical Science.

Die Vorüberlegung, dass höher gerankte Zeitschriften eher eine Data Policy haben als niedriger gerankte, konnte für die geringe Anzahl an Zeitschriften, die für jede Liste analysiert wurde, nicht bestätigt werden. Für sechs Listen wurden jeweils die Anzahl an Zeitschriften mit einer Data Policy bei den Zeitschriften 1–10 und 11–20 erhoben und verglichen (Tabelle 1). In zwei Listen gab es mehr Zeitschriften mit einer Data Policy bei den Zeitschriften 1–10, in vier Listen gab es jedoch mehr Zeitschriften mit einer Data Policy bei den Zeitschriften 11–20. Der Anteil an Zeitschriften mit einer Data Policy unter den ersten 20 Zeitschriften der Listen lag im Durchschnitt bei 61,7%. Bei der Zählung wurden Zeitschriften, bei denen keine Analyse möglich war, nicht berücksichtigt.

Stärke der Data Policies

Es wurden 534 Zeitschriften untersucht. Von diesen hatten insgesamt 346 eine Data Policy mit folgenden Abstufungen: 68 (12,7%) strong, 67 (12,5%) weak und 211 (39,5%) optional (siehe Abb. 4).

Die Zeitschriften wurden 17 Fachgebieten zugeordnet (siehe Abb. 5). Die Fachgebiete umfassten neben naturwissenschaftlichen auch sozial- und geisteswissenschaftliche Bereiche, z. B. Archaeology, Criminology, Law und Sociology and Political Sciences. Im Durchschnitt wurden 31 Zeitschriften pro Fachgebiet analysiert, mit einer Schwankungsbreite von 17 in Law bis 74 in Psychology. Letzteres Fachgebiet ist ein Spezialfall, da 71 Zeitschriften der *American Psychological Association* aufgenommen wurden. Aufgrund einheitlicher Verlagsrichtlinien galt für alle diese Zeitschriften dieselbe optional Data Policy.

Die Analyse ergab, dass es in jedem Fachgebiet Zeitschriften mit zumindest einer optional Data Policy gibt. In den drei Fachgebieten Archaeology, Criminology und Psychology gab es allerdings weder Zeitschriften mit einer strong noch mit einer weak Data Policy. Die drei Fachgebiete mit den meisten Zeitschriften ohne Data Policy waren Criminology (88,9%), Law (74,1%) und Astronomy (66,7%). Die drei Fachgebiete mit den meisten Zeitschriften mit einer Data Policy waren Psychology (100%), Chemistry (96,3%) und Biology (88,9%). Die drei Fachgebiete mit den meisten Zeitschriften mit einer strong Data Policy waren Biology (63,0%), Medical Sciences (40,7%) und Chemistry (29,6%). Diese Fachgebiete hatten (mit den Fachgebieten Earth Sciences und Statistics) auch den größten Anteil an Zeitschriften mit einer weak Data Policy: Chemistry (59,3%), Earth Sciences (29,6%), Biology (18,5%), Medical Sciences (18,5%), Statistics (18,5%). Abbildung 6 zeigt die Anzahl der strong und weak Data Policies in den einzelnen Fachgebieten.

In den Fachgebieten, in denen verhältnismäßig viele Zeitschriften ohne Data Policy (>30%) gefunden wurden, war auch tendenziell der Anteil an Zeitschriften mit einer optional Data Policy sehr hoch. Zu diesen Fachgebieten gehören Applied Sciences, Archaeology, Astronomy, Computer Science, Criminology, Law, Linguistics und Music. Drei Fachgebiete (Earth Sciences, Economics, Statistics) entsprechen allerdings nicht dieser Tendenz. Diese haben zwar auch zumindest 30% Zeitschriften ohne Data Policy, gleichzeitig ist der

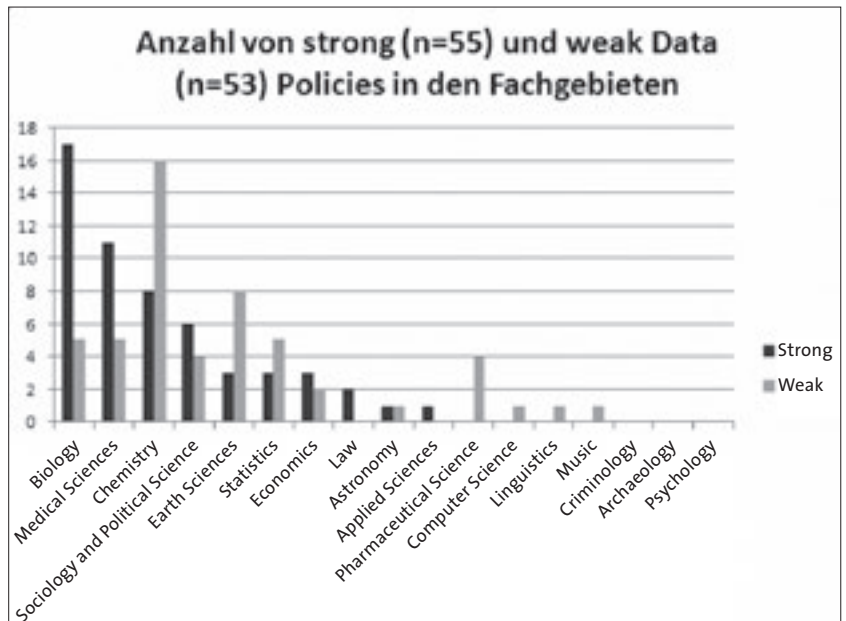


Abb. 6: Anzahl von strong und weak Data Policies in den Fachgebieten

Anteil an strong oder weak Policies aber verhältnismäßig hoch: Earth Sciences (40,7%), Economics (18,5%), Statistics (29,6%). Die prozentuale Verteilung der Stärken der Data Policies in den verschiedenen Fachgebieten ist in Abbildung 7 dargestellt.

Analyse-Ergebnisse

— Analyse der Keywords

Insgesamt wurden 346 Data Policies untersucht. Dabei wurden 59 verschiedene Keywords (inklusive der Sammelbegriffe ›specific‹ und ›replication material‹)

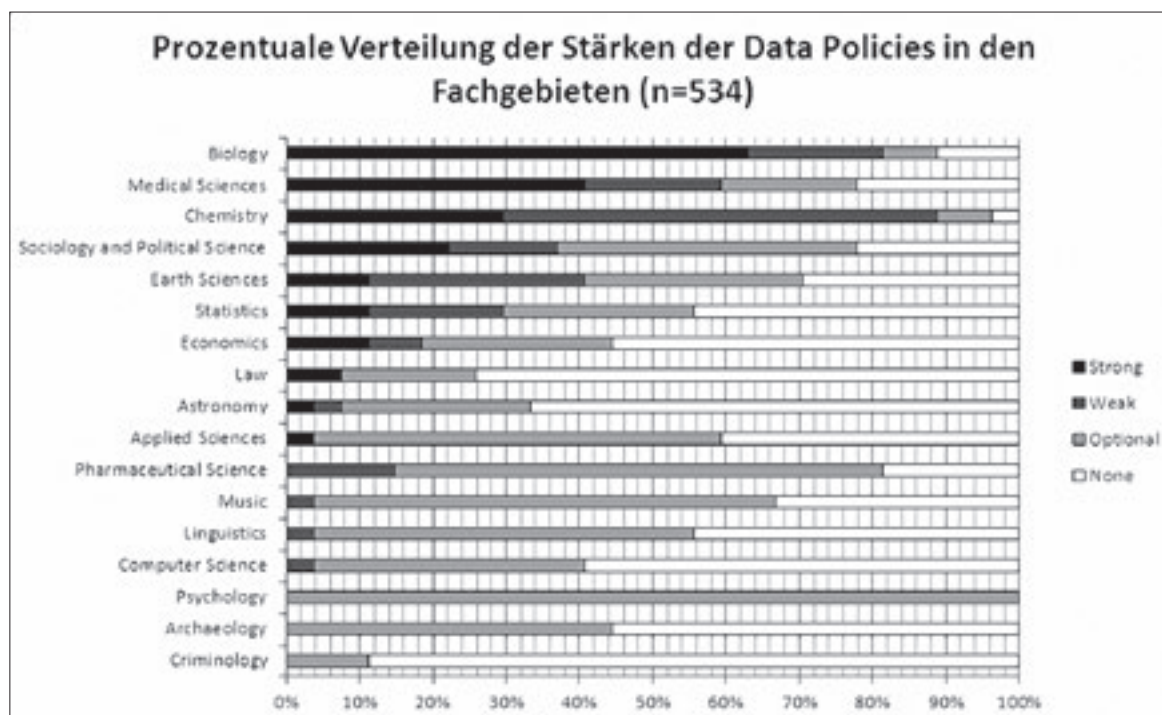


Abb. 7: Prozentuale Verteilung der Stärken der Data Policies in den Fachgebieten

Einteilung der Keywords in zwei Kategorien

identifiziert. In den strong Data Policies wurden insgesamt 37, in den weak Data Policies 27 und in den optional Data Policies 28 verschiedene Keywords vergeben. Die meisten Keywords setzten sich aus einem Nomen (z. B. data, data set, material, protocol, information, file) und einem Adjektiv (z. B. supplementary, additional, supporting, integral, underlying) zusammen. Die Keywords wurden im Hinblick darauf, ob sie einen inhaltlichen Bezug zu Forschungsdaten haben, in zwei Kategorien eingeteilt. Kategorie 1 enthält diejenigen Keywords, bei denen dies nicht der Fall war. Diese bezeichneten, grob gesagt, allgemeine Zusatzinformationen. Kategorie 2 enthält diejenigen Keywords mit einem inhaltlichen Bezug zu Forschungsdaten. Es war häufig der Fall, dass in einer Data Policy Keywords aus beiden Kategorien vorkamen. Die fünf häufigsten Keywords beider Kategorien waren folgende (geordnet in absteigender Reihenfolge):

- Kategorie 1: supplementary data, supplemental material, additional content, supplementary material, electronic supplementary information.
- Kategorie 2: integral content, specific, externally deposited data, replication material, raw data.

Die Anteile der Keywords aus Kategorie 2 nahm von strong zu optional ab: strong (79,2%), weak (49,2%), optional (22,5%). Dementsprechend waren die Anteile der Keywords aus Kategorie 1: strong (20,8%), weak (50,8%), optional (77,5%). Der Anteil an Keywords der Kategorie 2 in allen Data Policies war 41,6%. Es gab nur einen mit ›Forschungsdaten‹ synonymen Ausdruck, nämlich ›research material‹. Dieser kam insgesamt genau einmal vor. Es kamen mehrere Bezeichnungen für Rohdaten vor: ›raw data‹ kam fünfzehnmal vor, ›primary data‹ dreimal, die Ausdrücke ›raw supplementary data‹ und ›primary observational material‹ kamen jeweils einmal vor. Diese Ausdrücke wurden fast ausschließlich in strong und weak Data Policies verwendet, wo das Hinterlegen von Rohdaten gefordert wurde.

Im Folgenden sind die zehn häufigsten Keywords jeweils für optional, weak und strong Data Policies so-

wie für sämtliche Data Policies (strong, weak und optional) geordnet in absteigender Reihenfolge nach ihrer Häufigkeit angeführt. Die Keywords in Standardformatierung sind aus Kategorie 1 (Zusatzinformationen), diejenigen in Fettdruck sind aus Kategorie 2 (Forschungsdaten):

- optional: supplemental material, additional content, **integral content**, supplementary data, supplementary material, electronic supplementary material, supplemental online material, supplementary file, data supplement, supporting information;
- weak: supplementary data, **specific**, electronic supplementary information, **externally deposited data**, supplementary material, **raw data**, supplementary file, **replication material**, **relevant data**, **underlying data**;
- strong: **specific**, **replication material**, supporting data, supplementary data, supporting information, **underlying data**, **raw data**, electronic supplementary information, **externally deposited data**, supplementary information;
- strong, weak und optional: supplementary data, supplemental material, additional content, **integral content**, **specific**, supplementary material, electronic supplementary information, **externally deposited data**, electronic supplementary material, **replication material**.

Die folgende Keywords Cloud⁶ enthält alle identifizierten Keywords und soll einen zusätzlichen Eindruck davon vermitteln, welche Keywords besonders häufig vorkamen.

DISKUSSION

Suchstrategien

Es konnte mit keiner der Suchstrategien ein optimales Ergebnis erzielt werden. Der größte Nachteil war, dass die erzeugten Listen fast immer auch Zeitschriften enthielten, die keine Data Policy hatten. Daher musste auch für die Analyse von Zeitschriften ohne Data Policy Zeit aufgewendet werden. Der zweite Nachteil

Keywords Cloud

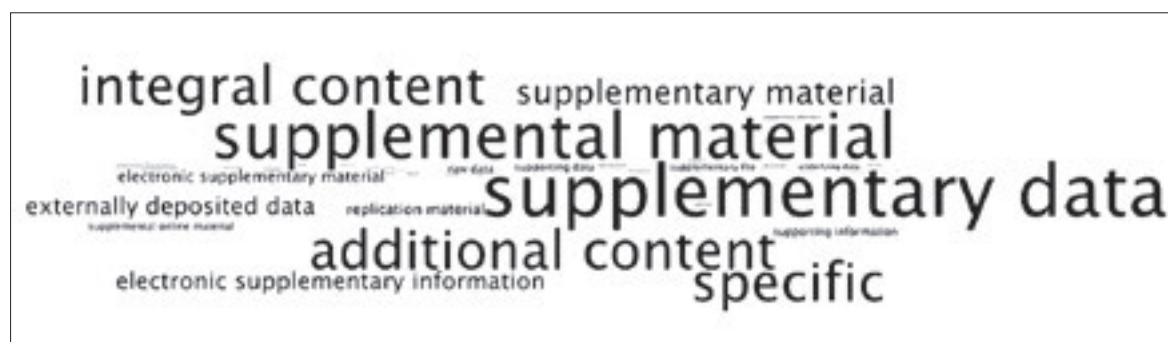


Abb. 8: Keywords Cloud

war, dass – von wenigen Repositorien (z. B. *Dryad*) abgesehen – keine direkte Verlinkung zur Website der Zeitschriften vorhanden war. Auch die Recherche dazu nahm Zeit in Anspruch. In einigen Fällen konnte die Website der Zeitschrift nicht ausfindig gemacht werden, weshalb keine Analyse möglich war. Im Folgenden wird auf die Vor- und Nachteile der verschiedenen Suchstrategien eingegangen.

Google Scholar

Google Scholar wurde nach der Testphase nicht mehr für die Suche eingesetzt. Der Grund dafür war, dass unter den Suchergebnissen regelmäßig nur sehr wenige relevante Treffer waren. Google Scholar eignete sich aber gut für die Suche nach Literatur über Data Policies. Außerdem wurden ein Repository (*Dryad*) und ein Fachverlag (*American Psychological Association*) über Google Scholar gefunden.

Bewertung:

- + Literatur über Data Policies
- sehr niedrige Precision

Repositorien

Die hohe Trefferquote und die große Anzahl an strong Data Policies in naturwissenschaftlichen Fachgebieten legen nahe, dass Repositorien eine gute Quelle für naturwissenschaftliche Zeitschriften mit einer strong oder weak Data Policy sind. Für diese Fachgebiete gibt es bereits viele Repositorien, die außerdem oft Verweise auf die entsprechenden Zeitschriften enthalten. In anderen Fachgebieten war die Trefferquote deutlich geringer. Ein großes Problem bei nicht-naturwissenschaftlichen Fachgebieten war, dass kein Repository gefunden werden konnte. Viele der gefundenen Repositorien waren außerdem für die Suchstrategie nicht geeignet, weil nur Rohdaten (statistische Messreihen, Bildmaterialien, Texte) enthalten waren oder kein Verweis auf Zeitschriften, in denen Ergebnisse zu den hinterlegten Daten publiziert worden waren, vorhanden war. Für Fachgebiete außerhalb der Naturwissenschaften sind Repositorien deshalb keine geeignete Quelle, um Zeitschriften mit einer Data Policy zu finden.

Bewertung:

- + sehr hohe Erfolgsquote in Naturwissenschaften
- wenig ergiebig in nicht naturwissenschaftlichen Fachgebieten
- Dublettengefahr, intellektuelle Kontrolle notwendig

Sonstiges: Eine alternative Quelle für diesen Ansatz wäre der *Data Citation Index*⁷ von Thomson Reuters. Dieser enthält Metadaten zu ausgewählten Repositorien und den darin hinterlegten Datensätzen sowie die Verweise auf die Zeitschriften, in denen auf die Da-

tensätze Bezug genommen wird. Zum Zeitpunkt der Recherche war der *Data Citation Index* an der Universität Wien allerdings noch nicht verfügbar. Eine spätere Überprüfung ergab, dass die Navigation von den Datensätzen zu den Zeitschriften sehr umständlich ist und daher eine Recherche im *Data Citation Index* den Rahmen dieses Projektes gesprengt hätte.

Zeitschriften-Rankings

SCImago Journal & Country Rank 2012 wurde von allen vier Quellen am meisten genutzt und stellte dementsprechend auch den größten Anteil (etwa die Hälfte) an den gefundenen Zeitschriften dar. Der Vorteil dieser Quelle bestand darin, dass auch Listen von Zeitschriften aus geisteswissenschaftlichen Fachgebieten verfügbar waren, die über die anderen Suchstrategien nicht oder nur mit großem Aufwand gefunden hätten werden können. Außerdem konnten die Zeitschriften-Rankings systematisch abgearbeitet werden. Für breitangelegte Analysen zu Data Policies, die auch geisteswissenschaftliche Fachgebiete umfassen, sind Zeitschriften-Rankings deshalb als Quelle gut geeignet. Der verhältnismäßig geringe Anteil an strong und weak Data Policies könnte dadurch zustande gekommen sein, dass fast ausschließlich Zeitschriften-Rankings von nicht-naturwissenschaftlichen Fachgebieten analysiert wurden. Vermutlich ist der Anteil an strong und weak Data Policies in naturwissenschaftlichen Zeitschriften-Rankings deutlich höher.

Die Ergebnisse von JoRD legen nahe, dass in den *Journal Citation Reports* ein Zusammenhang zwischen der Rankingposition und dem Vorhandensein einer Data Policy bei einer Zeitschrift besteht: Aus der Science Edition der *Journal Citation Reports* 2012 wurden 98 der top 100 Zeitschriften und 80 der bottom 100 analysiert. 92 (93,9 %) der top 100, aber nur 10 (12,5 %) der bottom 100 Zeitschriften hatten eine Data Policy. Aus der Social Sciences Edition der *Journal Citation Reports* 2012 wurden alle top 100 Zeitschriften und 93 der bottom 100 analysiert. 60 (60 %) der top 100, aber nur 5 (5,4 %) der bottom 100 Zeitschriften hatten eine Data Policy. Dieses Ergebnis konnte für die Zeitschriften-Rankings im *SCImago Journal & Country Rank* nicht überprüft werden, weil die Anzahl der untersuchten Zeitschriften pro Liste zu gering war. Das Ergebnis, dass vier von sechs Listen mehr Zeitschriften mit einer Data Policy unter den Zeitschriften 11–20 hatten als unter den Zeitschriften 1–10, kann aber auch nicht als Beleg dafür gewertet werden, dass dieser Zusammenhang nicht besteht, weil der untersuchte Bereich der Listen zu klein war. Dieses Ergebnis verdeutlicht aber, dass ein sehr hoher Platz in einem Zeitschriften-Ranking keine Garantie für eine Data Policy ist. Be-

Vor- und Nachteile verschiedener Suchstrategien

naturwissenschaftliche Zeitschriften

Rohdaten

achtlich ist trotzdem der hohe Anteil an Data Policies unter den Zeitschriften 1–20 der Listen. Dieser lag im Durchschnitt bei über 60 %. Die Hälfte der Zeitschriften hatte sogar einen Anteil von 80 %.

Bewertung:

- + einfach anwendbar
- + Zugang zu geisteswissenschaftlichen Zeitschriften
- die hohe Reihung einer Zeitschrift ist unabhängig vom Fachgebiet keine Garantie für das Vorhandensein einer Data Policy
- Dublettengefahr bei Rankings verwandter Fachgebiete

Sonstiges: Die *Journal Citation Reports* 2012 (die zum Zeitpunkt der Recherche an der Universität Wien aktuellste Ausgabe) wurde für die Recherche nicht verwendet, weil JoRD die *Journal Citation Reports* 2012 als Basis für seine Recherche verwendet und die bestgereihten Zeitschriften dieses Rankings bereits analysiert hatte. Ein Kopieren der Vorgangsweise von JoRD erschien deshalb nicht wünschenswert.

Websites von Fachverlagen

Die Websites von Verlagen war die Quelle mit der höchsten Trefferquote. Der Grund dafür war, dass die Zeitschriften eines Fachverlags nur dann alle untersucht wurden, wenn eine einheitliche Data Policy für alle Zeitschriften des Verlags vorhanden war. Die untersuchten Fachverlage gehörten oft zu einer wissenschaftlichen Gesellschaft (z. B. die *Royal Society of Chemistry*). Die Vorteile dieser Quelle waren, dass erstens die Zeitschriften in der Regel auf der Website als Liste vorhanden waren und so systematisch abgearbeitet werden konnten, und zweitens, dass diese Listen sehr wenige Zeitschriften ohne Data Policy enthielten. Der größte Nachteil war, dass das Finden eines geeigneten Verlages schwierig war. Jeder Treffer aus dieser Quelle stammt indirekt aus einer der drei anderen Quellen: Die Zeitschriften der *American Psychological Association* stammen zum Beispiel indirekt aus Google Scholar, die Zeitschriften der *Royal Society of Chemistry* aus einem Repositorium. Dadurch, dass die Zeitschriften aus dieser Quelle aus wenigen Fachverlagen stammen, ist diese Quelle für repräsentative Untersuchungen über die Häufigkeit von Data Policies in der gesamten Verlagslandschaft nicht geeignet.

Bewertung:

- + bei Vorliegen eines geeigneten Verlages effektiv und effizient
- Finden eines geeigneten Verlages schwierig
- einseitig, liefert keine repräsentativen Ergebnisse

Sonstiges: Die für Verlage genannten Überlegungen und Beobachtungen gelten genauso für wissenschaft-

liche Gesellschaften, die als Verleger von Zeitschriften auftreten.

Stärke der Data Policies

Der Anteil der Zeitschriften mit einer strong oder weak Data Policy war mit 25,3 % erheblich kleiner als derjenige der Zeitschriften mit einer optional oder keiner Data Policy (39,5 % + 35,2 %). Der Anteil an Zeitschriften ohne Data Policy ist in Wirklichkeit vermutlich deutlich höher als in der vorliegenden Menge von Zeitschriften, weil im Rahmen der Recherche bewusst nach Zeitschriften mit einer Data Policy gesucht wurde. Im Vergleich mit den Ergebnissen von JoRD ist der Anteil der Zeitschriften mit einer strong oder weak Data Policy gering (25 % vs. 45 %). Dafür erscheinen zwei Erklärungen plausibel:

- Der Anteil an naturwissenschaftlichen Zeitschriften bei JoRD ist verhältnismäßig groß (48 %), darunter sind die 98 bestgereihten Zeitschriften der Science Edition des *Journal Citation Reports* 2012, die 58 % aller Data Policies von JoRD ausmachen.
- JoRD verwendete zwei Kategorien von Data Policies (strong, weak) statt drei (strong, weak, optional). Möglicherweise gab es dadurch Unterschiede in der Kategorisierung von strong und weak Data Policies. Dafür spricht auch das Ergebnis, dass in diesem Projekt der Anteil an weak und strong Data Policies annähernd gleich war, während bei JoRD der Anteil an weak Data Policies deutlich höher war (85 %). Ein direkter Vergleich der Ergebnisse ist deshalb in diesem Fall nicht möglich.

Eine Besonderheit bei den Zeitschriften mit einer optional Data Policy war, dass diese überwiegend zu Großverlagen mit einer einheitlichen Data Policy für den Großteil ihrer Zeitschriften gehörten. Diese war dementsprechend allgemein und unverbindlich formuliert. Der Autor wird in einer solchen Data Policy auf die Möglichkeit hingewiesen, Publikationen mit Zusatzmaterialien zu ergänzen, wobei aber nicht explizit auf Forschungsdaten Bezug genommen wird. Solche Verlagsrichtlinien fanden sich z. B. bei Elsevier, Wiley-Blackwell und Springer. Das Vorhandensein von Verlagsrichtlinien schließt aber nicht aus, dass für manche Zeitschriften des Verlages individuelle strong oder weak Data Policies formuliert worden sind, welche von den allgemeinen Verlagsrichtlinien abweichen können. Die meisten Zeitschriften mit einer strong oder weak Data Policy waren aus naturwissenschaftlichen Fachgebieten (Biology, Chemistry, Medical Sciences). Dies bestätigt das Ergebnis von JoRD, wo Zeitschriften aus der Science Edition der *Journal Citation Reports* (mit einem hohen Impact Factor) am häufigsten eine strong oder weak Data Policy hatten.

Fehlen von Data Policies

wissenschaftliche
Gesellschaften und
Großverlage

Verlagsrichtlinien

In Einklang mit diesem Ergebnis waren in diesem Projekt die Fachgebiete mit den meisten strong oder weak Data Policies klassisch naturwissenschaftliche Fachgebiete (Biology, Chemistry, Medical Sciences). Es gab allerdings auch naturwissenschaftliche Fachgebiete, wo strong oder weak Data Policies deutlich schwächer vertreten waren (Astronomy, Pharmaceutical Science, Earth Sciences). Im Fachgebiet Astronomy hatten 66,7% der Zeitschriften keine Data Policy, 25,9% eine optional Data Policy und nur 7,4% hatten eine strong oder weak Data Policy. Das ist insofern bemerkenswert, weil in der Astronomie sehr große Mengen an Daten gesammelt werden (Sternkataloge, Oberflächenbilder, Atmosphärendaten) und es auch zahlreiche Fachrepositorien gibt, die das Hinterlegen von Daten ermöglichen: alleine in der *Registry of Research Data Repositories* sind 83 Repositorien diesem Fachgebiet zugeordnet. Dies legt nahe, dass es zwar bestimmte naturwissenschaftliche Fachgebiete gibt, in denen Data Policies weit verbreitet sind, dies aber nicht verallgemeinert werden darf. Das Ergebnis, dass es in geisteswissenschaftlichen Fachgebieten wie Criminology, Law, Linguistics oder Music deutlich weniger strong und weak Data Policies gab als in naturwissenschaftlichen Fachgebieten entspricht ebenfalls den Ergebnissen von JoRD. Eine mögliche Erklärung für dieses Ergebnis ist der große Anteil an theoretischen Arbeiten in den Geisteswissenschaften.

— Analyse der Keywords

Die Formulierungen in Data Policies, die sich auf die zu hinterlegenden Daten beziehen, waren oft sehr allgemein formuliert. Die große Anzahl der identifizierten Keywords legte nahe, dass sich bisher keine einheitlichen Formulierungen für Data Policies etabliert haben. Je strenger eine Data Policy formuliert war, desto eher fanden sich genaue Angaben, welche Daten zu hinterlegen sind. Das belegen folgende Ergebnisse:

- Der Sammelbegriff »specific« war bei strong Data Policies das häufigste und bei weak Data Policies das zweithäufigste Keyword, bei den optional Data Policies aber nicht unter den zehn häufigsten Keywords.
- Keywords der Kategorie 2 (Forschungsdaten) kamen sehr häufig in strong Data Policies vor, während Keywords der Kategorie 1 sehr häufig in optional Data Policies vorkamen. Bei den weak Data Policies war das Verhältnis der beiden Kategorien ausgewogen.
- Bezeichnungen für Rohdaten kamen fast ausschließlich in strong und weak Data Policies vor.

Das Ergebnis, dass Synonyme zu »Forschungsdaten« praktisch nicht vorkamen, könnte damit zu tun ha-

ben, dass es bislang keine allgemeingültige Definition dieses Terminus gibt und die Verwendung vermieden wird, um Missverständnisse zu vermeiden. Zeitschriften, die wollten, dass Forschungsdaten hinterlegt werden, gaben stattdessen häufig konkret an, welche Daten zu hinterlegen sind, oder verwendeten Formulierungen, die deutlich machten, dass die Daten zu hinterlegen sind, welche die präsentierten Ergebnisse vollständig nachvollziehbar und reproduzierbar machen. Auf diese Tendenz weist das Ergebnis hin, dass bei den strong Data Policies die beiden Sammelbegriffe »specific« und »replication material« die beiden häufigsten Keywords waren. Auch bei den weak Data Policies waren »specific« und »replication material« verhältnismäßig häufig (Platz 2 bzw. Platz 8).

AUSBLICK

Aufbauend auf den bisherigen Ergebnissen folgen weitere Analysen zu Data Policies. In einem ersten Schritt wurde inzwischen bereits untersucht, inwieweit der *Data Citation Index* als Datenquelle für Zeitschriften im Rahmen dieses Projektes dienen kann. Außerdem soll die Liste der gesammelten Kontaktdaten der Zeitschriftenredaktionen als Grundlage für eine spätere Befragung zu Data Policies dienen. Da es auch international noch wenige Arbeiten zum Thema Data Policies gibt, wären u. a. folgende weiterführende Untersuchungen von Interesse:

- eine ausführliche Analyse des *Data Citation Index* als potentielle Quelle für Zeitschriften mit einer Data Policy, und
- Untersuchungen dazu, ob in Zeitschriften-Rankings wie *SCImago Journal & Country Rank* ein direkter Zusammenhang zwischen der Rankingposition einer Zeitschrift und dem Vorhandensein einer Data Policy – insbesondere strong Data Policies – besteht.

Es wird interessant sein, zu beobachten, in welchem Umfang sich die Verbreitung von Data Policies bei wissenschaftlichen Zeitschriften verändern wird, und ob diese in den Geistes- und Sozialwissenschaften an Bedeutung gewinnen werden. Letztendlich weist diese Studie auf die Notwendigkeit hin, Informationen bezüglich Forschungsdaten und etwaiger zugehöriger Policies in den herkömmlichen Zeitschriften-Directories wie z. B. *Ulrichsweb* oder DOAJ zu inkludieren.

DANKSAGUNG

Dieses Projekt wurde im Rahmen von e-Infrastructures Austria (www.e-infrastructures.at/) und des Universitätslehrganges »Library and Information Studies« der Universität Wien durchgeführt. Der Name des Projektes war »Richtlinien wissenschaftlicher Zeitschriften für das Hinterlegen von Forschungsdaten«. Hiermit

keine allgemeingültige Definition

keine einheitlichen Formulierungen

möchten wir uns auch bei Frau Barbara Sánchez Solís für ihre Unterstützung und Anregungen bedanken.

Die Gesamtliste der untersuchten Zeitschriften ist im Web unter <http://zfbk.klostermann.de/data/policies.html> einsehbar.

LITERATUR

Bamkin, Marianne; Smith, Jane H.; Sturges, Paul; Hussain, Azhar; Hubbard, Bill: Publisher interest towards a role for journals in data-sharing. Extended Abstract: The findings of the JoRD Project. Centre for Research Communication, University of Nottingham. Verfügbar unter: http://elpub.architexturez.net/system/files/pdf/115_elpub2013.content.04273.pdf (Zugriff am: 27.05.2014).

Hubbard, Bill; Hussain, Azhar; Bamkin, Marianne; Smith, Jane H.; Sturges, Paul; Goldstein, Stephen: JoRD-Projektblog. Verfügbar unter: <http://jordproject.wordpress.com> (Zugriff am: 03.06.2014).

Ludwig, Jens: Zusammenfassung und Interpretation. In: Neuroth, Heike et al. (Hrsg.) (2012): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. 1. Aufl., Boizenburg: Verlag Werner Hülsbusch, S. 295–310. Verfügbar unter: <http://nestor.sub.uni-goettingen.de/bestandsaufnahme/>

McCain, Katherine W.: Mandating Sharing – Journal Policies in the Natural Sciences. In: Science Communication, Vol. 16, No 4 (1995), S. 403–431.

Oßwald, Achim; Scheffel, Regine; Neuroth, Heike: Langzeitarchivierung von Forschungsdaten. Einführende Überlegungen. In: Neuroth, Heike et al. (Hrsg.) (2012): Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme. 1. Aufl., Boizenburg: Verlag Werner Hülsbusch, S. 13–21. Verfügbar unter: <http://nestor.sub.uni-goettingen.de/bestandsaufnahme/>

Piowar, Heather A.; Chapman, Wendy W.: A review of journal policies for sharing research data. Department of Biomedical Informatics, University of Pittsburgh School of Medicine. Verfügbar unter: <http://preceedings.nature.com/documents/1700/version/1/html> (Zugriff am: 27.05.2014).

Sánchez Solís, Barbara: Factors for Enabling Sharing and Reuse of Research Data. A Study performed by NOAD Austria. Version 1.0. Verfügbar unter: <https://fedora.phaidra.univie.ac.at/fedora/get/o:343651/bdef:Content/get> (Zugriff am: 28.05.2014).

Vlaeminck, Sven; Siegert, Olaf: Welche Rolle spielen Forschungsdaten eigentlich für Fachzeitschriften? Eine Analyse mit Fokus auf die Wirtschaftswissenschaften. RatSWD Working Paper Series. Verfüg-

bar unter: http://juwel.fz-juelich.de:8080/dspace/bitstream/2128/4699/1/Bibliothek_21.pdf (2012).

Wicherts, Jelte M.; Bakker, Marjan; Molenaar, Dylan: Willingness to Share Research Data Is Related to the Strength of the Evidence and the Quality of Reporting of Statistical Results. In: PLoS ONE, No. 11, 2011 (Vol. 6) pp. 1–7.

¹ www.e-infrastructures.at/; <https://rd-alliance.org/>; <http://ec.europa.eu/programmes/horizon2020/>; www.dfg.de/foerderung/programme/infrastruktur/lis/lis_foerderangebote/forschungsdaten/

² Aus Gründen der besseren Lesbarkeit wird auf die gleichzeitige Verwendung männlicher und weiblicher Sprachformen verzichtet. Sämtliche Personenbezeichnungen gelten gleichwohl für beiderlei Geschlecht.

³ Das Projekt wurde im Rahmen der »e-infrastructure Austria« (www.e-infrastructures.at/) und des Universitätslehrganges Library and Information Studies der Universität Wien durchgeführt.

⁴ <http://jordproject.wordpress.com/>

⁵ Eine ähnliche, allerdings inhaltlich anders motivierte Kategorisierung von Data Policies hat Katherine W. McCain in dem Aufsatz *Mandating Sharing* vorgeschlagen (vgl. McCain, 1995, S. 408). Sie teilt Data Policies anhand der »penalties for noncompliance« in vier Kategorien ein, wovon die ersten drei eine vergleichbare Einteilung der Zeitschriften zur Folge haben wie die Kategorien strong, weak und optional: (1) Paper not published (stated or strongly implied); (2) Paper not published (weaker language); (3) Nothing specified; (4) Denied future publication access. Das Kategorisierungsmerkmal bei McCain war also, wie deutlich das Nicht-Hinterlegen von Daten laut Data Policy mit einer Sanktion (Nichtveröffentlichung) verbunden war. Inhaltlich entsprechen diese Kategorien deshalb auch der Verpflichtung, der Aufforderung und der Möglichkeit Daten zu hinterlegen.

⁶ Die Grafik wurde mit Wordle erstellt.

⁷ <http://thomsonreuters.com/data-citation-index/>

DIE VERFASSER

Benedikt Blahous, BA, Absolvent des Universitätslehrgangs »Library and Information Studies« (Grundlehrgang) an der Universität Wien

Dr. Juan Gorraiz, Universität Wien, Bibliotheks- und Archivwesen, Team Bibliometrie, Boltzmann-gasse 5, A-1090 Wien, Tel.: +43-1-4277-276 09, E-Mail: juan.gorraiz@univie.ac.at

Dr. Christian Gumpenberger, Universität Wien, Bibliotheks- und Archivwesen, Team Bibliometrie, Boltzmann-gasse 5, A-1090 Wien, Tel.: +43-1-4277-276 19, E-Mail: christian.gumpenberger@univie.ac.at

Oliver Lehner, Technische Universität Wien, Universitätsbibliothek, IT Services, Resselgasse 4, A-1040 Wien, Tel.: +43-1-58801-44171, E-Mail: oliver.lehner@tuwien.ac.at

Mag. Bettina Stein, Absolventin des Universitätslehrgangs »Library and Information Studies« (Grundlehrgang) an der Universität Wien

Mag. Ursula Ulrych, Absolventin des Universitätslehrgangs »Library and Information Studies« (Grundlehrgang) an der Universität Wien