

Dallmeier-Tiessen, Sünje

Working Paper

Strategien bei der Veröffentlichung von Forschungsdaten

RatSWD Working Paper, No. 173

Provided in Cooperation with:
German Data Forum (RatSWD)

Suggested Citation: Dallmeier-Tiessen, Sünje (2011) : Strategien bei der Veröffentlichung von Forschungsdaten, RatSWD Working Paper, No. 173, Rat für Sozial- und Wirtschaftsdaten (RatSWD), Berlin

This Version is available at:
<http://hdl.handle.net/10419/75349>

Standard-Nutzungsbedingungen:

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

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

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

Terms of use:

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

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

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



Rat für Sozial- und
Wirtschaftsdaten (RatSWD)

www.ratswd.de

RatSWD

Working Paper Series

Working Paper

Nr. 173

Strategien bei der Veröffentlichung
von Forschungsdaten

Sünje Dallmeier-Tiessen

März 2011

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

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

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

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

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

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

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

Herausgeber der RatSWD Working Paper Series:

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

Geschäftsführer des RatSWD (Denis Huschka)

Strategien bei der Veröffentlichung von Forschungsdaten

Sünje Dallmeier-Tiessen*

Zusammenfassung

Forschungsdaten liegen in Abhängigkeit der Disziplinen in vielfältigen Formen und Formaten vor. Sie sind in allen Disziplinen Teil des wissenschaftlichen Erkenntnisprozesses. Als digitales Informationsobjekt sind sie komplex und bislang wenig untersucht. Mit den Möglichkeiten neuer Informationstechnologien werden in den letzten Jahren neue Wege in der Publikation von Forschungsdaten beschritten. Mit Blick auf die Naturwissenschaften werden im Folgenden drei Publikationsmodelle beschrieben: Die Veröffentlichung von Forschungsdaten als eigenständiges Objekt in einem Forschungsdatenrepositorium, die Veröffentlichung von Forschungsdaten mit textueller Dokumentation und die Veröffentlichung von Forschungsdaten als Anreicherung einer interpretativen Text-Publikation.

* gegenwärtig am CERN und Promovierende der Humboldt Universität zu Berlin (IBI), Kontakt: sunje.dallmeier-tiessen@cern.ch

1. Einführung

Das Verständnis des Begriffes Forschungsdaten variiert je nach Disziplin. Möchte man sich dem Begriff disziplinübergreifend annähern, so lässt sich dies z.B. über die Positionierung der Forschungsdaten im wissenschaftlichen Arbeitsprozess versuchen. Forschungsdaten sind der zentrale Gegenstand des wissenschaftlichen Erkenntnisprozesses. Ausgehend von bereits veröffentlichten Ergebnissen der Community werden sie ausgewertet und eingeordnet. Im Rahmen des Publikationsprozesses werden sie anschließend beschrieben. Diese Beschreibung in textueller Form stellt traditionell eine Interpretation und Diskussion der Daten dar. Dabei sind je nach Disziplin unterschiedliche Kulturen etabliert, die auf den heterogenen Formen und Formaten der Daten basieren. Während in den Naturwissenschaften beispielsweise oftmals ein Untersuchungsgegenstand durch unterschiedliche Methoden erhoben und ausgewertet wird, wird in den Geisteswissenschaften häufig auf bereits publizierte Untersuchungsgegenstände wie historische Text- und Bildmaterialien Bezug genommen, die traditionell unter dem Begriff „Quellenmaterial“ gefasst werden.

Am Beispiel der Biologie wird die Heterogenität auch innerhalb der Disziplinen deutlich: In dieser Disziplin können z.B. numerische Daten von Laborexperimenten, audiovisuelle Objekte wie Fotos oder Videosequenzen von Tierbeobachtungen, oder komplexe Modelle von Simulationen (z.B. von Umweltveränderungen) als Forschungsdaten bezeichnet werden.

Da eine detaillierte interdisziplinäre Definition des Begriffs Forschungsdaten schwerfällt, werden für die folgende konzeptionelle Betrachtung der Publikation von wissenschaftlichen Daten, Forschungsdaten als digitale Informationsobjekte betrachtet, die während des Forschungsprozesses entstehen und als Grundlage einer interpretativen wissenschaftlichen Textpublikation dienen.

Mit der rasanten Entwicklung der Informationstechnologien haben sich in den letzten Jahren neue Wege im Umgang mit wissenschaftlichen Daten eröffnet, die nach dem Verständnis von Jim Gray zu einem neuen Paradigma des wissenschaftlichen Arbeitens geführt haben:

„The world of science has changed, and there is no question about this. The new model is for the data to be captured by instruments or generated by simulations before being processed by software

and for the resulting information or knowledge to be stored in computers. Scientists only get to look at their data fairly late in this pipeline. The techniques and technologies for such data-intensive science are so different that it is worth distinguishing data-intensive science from computational science as a new, fourth paradigm for scientific exploration.“ (Hey et al. 2009)

Mit diesem „fourth paradigm“ entstehen neue Wege im Umgang mit wissenschaftlichen Daten.

2. Diskussion um den zeitgemäßen Umgang mit Forschungsdaten

In Deutschland wird der Umgang mit Forschungsdaten vornehmlich in Bezug auf die Nachprüfbarkeit der Daten geregelt. Ein zentraler Punkt sind dabei die von der Deutschen Forschungsgemeinschaft (DFG) 1998 veröffentlichten „Vorschläge zur Sicherung guter wissenschaftlicher Praxis“. Diese sehen vor, dass „Primärdaten als Grundlagen für Veröffentlichungen [...] auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, für zehn Jahre aufbewahrt werden [sollen].“ (DFG 1998). Der Aspekt der Nachnutzung gewinnt aktuell an Relevanz. 2010 hat die DFG diesen Aspekt im „Leitfaden für Antragsteller“ verankert. In diesem heißt es: „Wenn aus Projektmitteln systematisch (Mess-)Daten erhoben werden, die für die Nachnutzung geeignet sind, legen Sie bitte dar, welche Maßnahmen ergriffen wurden bzw. während der Laufzeit des Projektes getroffen werden, um die Daten nachhaltig zu sichern und ggf. für eine erneute Nutzung bereit zu stellen.“ (DFG 2010) Andere Förderer sind in ihren Anforderungen noch konkreter, so erwartet die National Science Foundation (NSF) einen Datenmanagementplan von Antragstellern (NSF 2011).

Auf internationaler Ebene wurden in den letzten Jahren zahlreiche Empfehlungen und Vorgaben zum Umgang mit Forschungsdaten verabschiedet. Zentral sind hier die 2007 von der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) publizierte „Principles and Guidelines for Access to Research Data from Public Funding“. In den Disziplinen variiert die Bereitschaft des „data sharings“ deutlich. In der Genomforschung wurden 1996 die „Bermuda Principles“ im Rahmen des Humangenomprojektes formuliert (z.B. Marshall 2001), nach denen Gensequenzen frei zugänglich gemacht werden sollen. Auch Fachgesellschaften wie z.B. die American Geophysical Union (AGU, 1996) haben Richtlinien zum Umgang mit Forschungsdaten formuliert. Die AGU-Richtlinie

bezieht sich auf Daten, die Grundlage einer Publikation sind, die in einem Organ der Fachgesellschaft veröffentlicht werden. Darüber hinaus haben auch Journale, wie z.B. PLoS ONE, in ihren Veröffentlichungsrichtlinien Aussagen zu dem Thema formuliert. PLoS ONE fordert unter dem Punkt „Sharing of Materials, Methods, and Data“ die Zugänglichkeit der Daten, die Basis einer Publikation sind (PloS 2011).

3. Publikationsmodelle

Auf Basis der beschriebenen Diskussion und der Annäherung an den Begriff „Forschungsdaten“ können diese Daten als zentraler Baustein des wissenschaftlichen Diskurses betrachtet werden.

Ausgehend von der Zugänglichkeit der Daten können unterschiedliche Aspekte einer wissenschaftlichen Nachnutzung herausgestellt werden, die auch auf die Publikationsstrategie der Daten wirken. Dabei sind die folgenden Ansätze zu unterscheiden:

- Die Publikation von Forschungsdaten um die Nachprüfung zu ermöglichen.
- Die Publikation von Forschungsdaten um die Nachnutzung zu ermöglichen.

Der häufig anzutreffende Begriff des „data sharing“ spezifiziert, dass eine Publikation von Forschungsdaten zum Zwecke der Nachnutzung durch Dritte erfolgt. Nach Borgman (2010) sprechen folgende Gründe für das „sharing“ wissenschaftlicher Daten: „to make the results of publicly funded data available to the public, to enable others to ask questions of extant data, to advance the state of science and to reproduce research“. Die beiden erstgenannten Argumente sind aus der Perspektive der Nachnutzung zu betrachten, deren Anspruch durch öffentliche Interessen gesteuert ist. Die beiden letztgenannten Argumente sind aus Sicht der Datenproduzenten interessant und greifen das Interesse der jeweiligen wissenschaftlichen Community auf (nach Borgman 2010).

Zentraler Akteur bei der Publikation von Forschungsdaten ist der Wissenschaftler oder das Team, das die Daten erhebt und interpretiert. Wird eine Veröffentlichung der Daten angestrebt, so bedarf es häufig einer erweiterten textuellen Aufbereitung der Daten von Seiten der Wissenschaftler, die im Folgenden unter dem Begriff „Dokumentation“ gefasst wird. In der Doku-

mentation werden die Kontextinformationen bereitgestellt, um Personen, die an der ursprünglichen Datenproduktion nicht beteiligt waren, eine Nachnutzung der Daten zu ermöglichen.

Da es in vielen Disziplinen bisher keine etablierte Kultur der Veröffentlichung wissenschaftlicher Daten gibt, ist der Aufwand einer nachnutzbaren oder nachprüfaren Veröffentlichung der Daten eine aufwendige und zusätzliche Tätigkeit, die über den „normalen“ Wissenschaftsalltag hinausgeht. Bei der Publikation der Daten müssen z.B., je nach Disziplin, rechtliche und ethische Rahmenbedingungen berücksichtigt werden. So wirken beispielsweise in den Sozialwissenschaften datenschutzrechtliche Aspekte auf die Publikationsmöglichkeiten und -strategien. Mit Blick auf die kompetitive Forschungslandschaft mag es dem einzelnen Wissenschaftler außerdem notwendig erscheinen, die Veröffentlichung von Daten nur eingeschränkt oder zeitverzögert anzustreben.

Ziel der folgenden übergreifenden Darstellung ist es, einen Überblick auf die gegenwärtige Landschaft der Forschungsdatenpublikation zu schaffen und damit eine Grundlage für zukünftige Diskussionen zu bieten. Hierbei soll vor allem herausgestellt werden, wie die wissenschaftlichen Daten in den traditionellen Publikationsprozess eingebunden sind.

Mit Blick auf die Naturwissenschaften können drei Publikationsmodelle unterschieden werden.

- Die Veröffentlichung von Forschungsdaten als eigenständiges Objekt in einem Datenrepositorium
- Die Veröffentlichung von Forschungsdaten mit einer textuellen Dokumentation
- Die Veröffentlichung von Forschungsdaten als Anreicherung einer interpretativen Text-Publikation.

Darüber hinaus gibt es weitere individuelle Lösungen, wie z.B. die individuelle Bereitstellung auf der eigenen Website. Die unterschiedlichen Modelle ziehen unter Umständen unterschiedliche Zeitpunkte der Forschungsdatenpublikation nach sich, insbesondere in Bezug auf den interpretativen Artikel. Dieser Aspekt soll hier nicht im Detail ausgeführt werden, er kann jedoch bei der Wahl des Publikationsmodells entscheidend sein.

3.1 Eigenständige Veröffentlichung im Forschungsdaten-Repository

Ein Publikationsmodell ist die Veröffentlichung von wissenschaftlichen Daten in einem Forschungsdaten-Repository. Das Repository ermöglicht es, Forschungsdaten zeitlich und räumlich unabhängig von einer interpretativen Veröffentlichung zu veröffentlichen. Die Daten werden als eigenständiges Objekt in einem Datenrepository abgelegt. Je nach Disziplin kann dies z.B. nach einer automatisierten Prozessierung aus einem „System“ heraus erfolgen. Dieses Publikationsmodell ist in vielen naturwissenschaftlichen Disziplinen verbreitet. Je nach Disziplin und Organisationsstruktur in den Disziplinen variieren die Anforderungen an die Standardisierung und Bereitstellung. Dieses Modell kann auch als Baustein verstanden werden, welches – Standardisierung und Interoperabilität vorausgesetzt – mit anderen Arbeitsabläufen und Produkten der Wissenschaft kombiniert werden kann, siehe dazu auch die Modelle in 3.2 und 3.3. Exemplarisch seien hier vier Repositorien genannt, die die Diversität und die Möglichkeiten dieses Modells widerspiegeln:

- PANGAEA (<http://www.pangaea.de>): In diesem Datenrepository werden geowissenschaftliche Forschungsdaten veröffentlicht. Dank der Nutzung des Digital Object Identifier (DOI) wird eine persistente Adressierung der Daten ermöglicht. Komplexere Forschungsdatensätze können als „Collection“ publiziert werden.
- GenBank (<http://www.ncbi.nlm.nih.gov/genbank>): In dieser Datenbank werden Forschungsdaten nach den disziplinären Regeln der Genforschung publiziert. Eine „accession number“ dient zur persistenten Adressierung. Nach Vorgaben der Editorial-Policies vieler lebenswissenschaftlicher Zeitschriften muss diese bei einer Interpretation der Daten in einem Aufsatz angegeben werden.
- Integrated Ocean Observing System (<http://www.ioos.gov>): Hier werden automatisch erfasste Messdaten der marinen Umweltforschung automatisiert aus dem System heraus veröffentlicht.
- Dryad (<http://datadryad.org>): Dieses Datenrepository der Biologie vergibt DOIs für die veröffentlichten Datensätze. Dryad konzentriert sich auf Datensätze, die Grundlage einer interpretativen Publikation sind.

3.2 Veröffentlichung mit einer textueller Dokumentation

In Hinblick auf die Nachnutzung und Nachvollziehbarkeit von Forschungsdaten benötigen diese oftmals eine erweiterte Dokumentation, die zwar keine Interpretation liefert, aber einen Kontext zur Erhebung der Daten liefert und Interessierten eine Nachnutzung ermöglicht. Vor diesem Hintergrund werden neuere Konzepte entwickelt. So werden z.B. Forschungsdaten mit einer begutachteten Dokumentation publiziert.

Im Folgenden sollen zwei Beispiele dieser Publikationsstrategie vorgestellt werden, bei denen Forschungsdaten mit einer nicht-interpretativen Dokumentation angereichert werden (siehe Abbildung 1).

Beispiel: Earth System Science Data (ESSD)

Die Open-Access-Zeitschrift Earth System Science Data (ESSD) widmet sich der erweiterten und qualitätsgesicherten Dokumentation von geowissenschaftlichen Forschungsdaten. Die Zeitschrift wird von Copernicus Publications verlegt und nutzt ein zweistufiges Peer-Review-Verfahren unter der Nutzung des „Public Peer-Reviews“ (Pöschl 2010), um einen Datensatz und dessen Dokumentation zu begutachten und zu veröffentlichen.

Der Publikationsprozess in ESSD ist wie folgt organisiert: Ein Publizierender verfasst eine Dokumentation eines Datensatzes. In dieser beschreibt er beispielsweise die verwendeten Instrumente und angewandten Verfahren bei der Bearbeitung der Daten. Diese Beschreibung, die wie ein klassischer Aufsatz gestaltet wird, reicht er dann unter Nennung eines Persistent Identifier des Datensatzes bei der Zeitschrift ein. Darüber hinaus veröffentlicht er die beschriebenen Daten auf einem frei zugänglichen Daten-Repository. Nach einem erfolgreichen Begutachtungsprozess wird dann die Datendokumentation publiziert. Nach diesem Ansatz sind Datenrepository und Datenjournal deutlich getrennt. Das Ambiente der Zeitschrift ESSD erinnert stark an die Journale, die „interpretative“ Artikel publizieren.

Das Datenjournal ist unabhängig von Repositorien, hat jedoch eigene Kriterien für die Akzeptanz der Repositorienwahl festgelegt (Pfeiffenberger 2011), welche sich nach den gegenwärtigen Standards für vertrauenswürdige Repositorien (z.B. von NESTOR 2009) richten.

Für die Wissenschaftler ist der Prozess der Publikation ein bekannter, obwohl das zentrale Publikationsobjekt Forschungsdaten und deren Dokumentation in diesem Rahmen neu ist. Der zusätzliche Aufwand einer Einreichung der Publikationsobjekte an zwei Publikationsplattformen (Zeitschrift und Datenrepositorium) wird für den Wissenschaftler durch den Erhalt eines zitierfähigen Artikels in einem bekannten Journalformat kompensiert.

Beispiel: Overlay Journal Infrastructure Meteorological Sciences (OJIMS)

Unter dem Begriff „Overlay Journal“ fasst man Publikationsmodelle zusammen, welche auf Textpublikationen zurückgreifen, die auf verschiedenen Open-Access-Plattformen publiziert sind. Das Konzept beinhaltet eine thematische Aggregation, meist mit einem weiteren Service kombiniert, z.B. einem Begutachtungs- oder Kommunikationsprozess. Das Konzept kann auch auf die Publikation von Forschungsdaten übertragen werden. Das Overlay Journal Infrastructure Meteorological Science (OJIMS) wendet dieses für meteorologische Daten an. Die meteorologischen Daten werden auf einem Forschungsdaten-Repositorium gespeichert, die Dokumentation in einem Dokumentenrepositorium. Beide Objekte werden anschließend von Fachwissenschaftlern begutachtet. Auf Basis dieser qualitätsgesicherten Publikation können dann Interpretationen der Daten vorgenommen werden.

3.3 (Interpretative) Publikation plus Forschungsdaten

Traditionell werden wissenschaftliche Ergebnisse in einer Textpublikation beschrieben und diskutiert. Über Disziplingrenzen hinweg lässt sich dabei seit Jahrhunderten eine ähnliche Artikelstruktur verfolgen, die auf die Darstellung eines Ergebnisses fixiert ist. Neben dem Text dienen Tabellen und Abbildungen der Beschreibung des dargestellten Sachverhaltes. Zusätzliche Materialien werden oft im Anhang dargestellt (siehe Abbildung 2). Mit der Digitalisierung der Wissenschaftskommunikation können diese Materialien (z.B. Messreihen) nutzbar integriert und publiziert werden. Die Publikation solcher Daten, die insbesondere zur Unterstützung einer interpretativen Publikation dienen, wird heute insbesondere auf den folgenden beiden Wegen umgesetzt:

- a) Die Daten werden auf der gleichen Plattform wie auch die Textpublikation veröffentlicht, aber nicht individuell persistent und standardisiert adressiert. Dies kann auf einer Verlags- oder auch Repositorienplattform sein, welche vorwiegend für Textpublikation ausgelegt ist. Die Daten sind somit Bestandteil des Artikels.
- b) In einigen Disziplinen werden Forschungsdaten als eigenständiges Objekt, das referenzierbar ist, auf der Publikationsplattform des textuellen Artikels gespeichert. Dank der Referenzierbarkeit, die über eine persistente Adressierung gegeben ist, kann der Datensatz zitiert werden. In dem interpretativen Artikel wird dann diese Identifizierung angegeben. Der Datensatz kann in diesem Modell auch in einem externen Datenrepositorium (siehe 3.1) zeitlich parallel zu dem interpretativen Artikel publiziert werden. Dies bedeutet für den Wissenschaftler, dass dieser den Forschungsdatensatz und den interpretativen Artikel an zwei unterschiedlichen Orten einreicht. Eine koordinierte Zusammenarbeit der involvierten Infrastruktureinrichtungen, zumeist Verlage und Datenrepositorien, wie sie in 3.1 beschrieben worden sind, ist dabei wichtig, um zusätzlichen Aufwand auf Seiten der Wissenschaftler zu vermeiden und um den Anforderungen einer vertrauenswürdigen Langzeitarchivierung zu entsprechen. In der Nachnutzung können Textpublikation und Forschungsdaten individuell genutzt und zitiert werden (siehe z.B. die Zitierempfehlung des Datenrepositoriums Dryad). Kooperationen zwischen den Publikationsagenten Verlag/Journal und Datenrepositorium lassen sich bereits in einigen Disziplinen beobachten. Es sollen hier zwei Beispiele vorgestellt werden.
- PANGAEA und Elsevier: Der Verlag Elsevier und das Datenrepositorium PANGAEA sind eine Kooperation eingegangen (Elsevier 2010), die erlaubt, dass Leser von Artikeln auf Science Direct direkten Zugang zu Forschungsdaten haben, die einem geowissenschaftlichen Artikel zugeordnet sind und auf PANGAEA abgelegt wurden. Der Link zu dem Datensatz ist prominent auf der Artikelseite platziert und führt den Leser direkt zum Datensatz.

- Dryad und verschiedene Verlage/Journale: Das Datenrepositorium Dryad kooperiert mit verschiedenen Verlagen und Journalen, insbesondere aus den biologischen Disziplinen. Das Dryad Konsortium hat eine „Joint Data Archiving Policy“ entwickelt, welche die Ablage von Forschungsdaten für Einreichungen in den Partnerjournalen regelt. Es wird explizit von Forschungsdaten gesprochen, die Grundlage der Interpretation im Artikel sind. Zitat: „data supporting the results in the paper“ (Datadryad 2010).

Einen interessanten Überblick über unterschiedliche disziplinspezifische Ausprägungen dieses Modells sind in den geförderten SURF-Projekten zu finden, welche an der disziplinspezifischen Umsetzung der so genannten „enhanced publications“ arbeiten (SURF 2010). SURF definiert „enhanced publications“ als „a publication – usually a text – that has been enhanced with additional material [...] The supplementary material may consist, for example, of research data, illustrative images, metadata sets, or post-publication data such as comments or rankings.“ Sechs disziplinspezifische Projekte werden von SURF seit Januar 2011 gefördert, u.a. aus der Ökonomie und der Linguistik.

4. Fazit

Die Entwicklung von Publikationsmodellen für Forschungsdaten findet sich aktuell in einer dynamischen Entwicklung, was eine Typisierung der unterschiedlichen Ansätze schwierig macht.

In dieser Übersicht wurden drei Modelle der Publikation von Forschungsdaten beschrieben. Im erstgenannten Modell werden Forschungsdaten zitierfähig auf einem Repostorium veröffentlicht. Im zweiten Modell werden Forschungsdaten mit einer „erweiterten“ Dokumentation einem Begutachtungsprozess unterzogen. Zusätzliche und qualitätsgesicherte Kontextinformationen erleichtern dann die Nachnutzung. Ein bereits in zahlreichen Disziplinen umgesetztes Modell ist die Forschungsdatenpublikation als Anhang zu einem interpretativen Artikel. Hierbei kann der Datensatz z.B. in einem externen Datenrepositorium abgelegt werden und dann in dem Artikel referenziert werden.

Alle Modelle setzen bestimmte Rahmenbedingungen in Bezug auf die genutzten Infrastrukturen voraus, insbesondere in Bezug auf eine vertrauenswürdige Publikationsumgebung, die beispielsweise Standards der Langzeitarchivierung berücksichtigt.

Die weitere Entwicklung der Publikationsmodelle wird stark von den Publikationskulturen der einzelnen Disziplinen abhängig sein. Aufgrund der hohen disziplinspezifischen Charakteristika der Forschungsdaten sind angepasste Lösungen, die den Anforderungen der Disziplinen entsprechen, nötig. Der Kooperation zwischen Wissenschaft, Infrastruktureinrichtungen und Verlagen kommt dabei eine entscheidende Rolle zu.

Literatur

Datenjournal:

Earth System Science Data (ESSD), published by Copernicus Publications:
<http://earth-system-science-data.net/>

Overlayjournal OJIMS:

<http://proj.badc.rl.ac.uk/ojims>

- AGU Publications Committee: Policy on Referencing Data in and Archiving Data for AGU Publications: 1993 [Revised March 1994, December 1995, October 1996], URL: http://www.agu.org/pubs/authors/policies/data_policy.shtml, abgerufen am 22.02.2011
- Bermuda Principles: Policies on Release of Human Genomic Sequence Data. URL: http://www.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml, abgerufen am 22.02.2011
- Borgman 2010; Research Data; Who will share what, with whom, when, and why? China-North American Library Conference, 2010, URL: <http://works.bepress.com/borgman/238/>, abgerufen am 22.02.2011
- Dallmeier-Tiessen, S. ; Pfeiffenberger, H.: Peer Reviewed Data Publication in Earth System Sciences. In: Puschmann, C. ; Stein, D. (Hrsg.) Towards Open Access Scholarship. Düsseldorf : Düsseldorf University Press, 2008, S. 77-84
- DataDryad: Joint Data Archiving Policy (JDAP), 2010, URL: <http://datadryad.org/jdap>, abgerufen am 22.02.2011
- Deutsche Forschungsgemeinschaft: Vorschläge zur Sicherung guter wissenschaftlicher Praxis : Empfehlungen der Kommission "Selbstkontrolle in der Wissenschaft" : Denkschrift. Weinheim : Wiley-VCH, 1998 – ISBN 9783527272129
- Deutsche Forschungsgemeinschaft. Merkblatt fuer Antraege auf Sachbeihilfe mit Leitfaden fuer die Antragsstellung. Bonn: Deutsche Forschungsgemeinschaft 2010. URL: http://www.dfg.de/download/programme/sachbeihilfe/antragstellung/1_02/1_02.pdf, abgerufen am 18.02.2011
- Elsevier: "Elsevier and PANGAEA Take Next Step in Connecting Research Articles to Data", 2010, URL: http://www.elsevier.com/wps/find/authored_newsitem.cws_home/companynews05_01616, abgerufen am 22.02.2011
- Hey, T. ; Tansley, S. ; Tolle, K.: Jim Gray on eScience: a transformed scientific method. In: Tansley, S. ; Tolle, K. (Hrsg.) The Fourth Paradigm: Data-Intensive Scientific Discovery. Redmond, Washington : Microsoft Research, 2009 – ISBN 978-0-9825442-0-4, S. XVII-XXXI. URL: <http://www.fourthparadigm.org>
- Marshall, E.: Bermuda Rules: Community Spirit, With Teeth. *Science* 16 February 2001: Vol. 291 no. 5507 p. 1192 DOI: 10.1126/science.291.5507.1192
- National Science Foundation: Data Management Plan in "Grant Proposal Guide (GPG) Chapter II.C.2.j", URL: http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp, 2011, abgerufen am 22.02.2011
- Nestor Kompetenznetzwerk: nestor-Kriterien - Kriterienkatalog vertrauenswürdige digitale Langzeitarchive, Version II / hrsg. von der nestor-Arbeitsgruppe Vertrauenswürdige Archive - Zertifizierung. Frankfurt am Main : nestor c/o Deutsche Nationalbibliothek, 2009. - 40 S.
- Organisation for Economic Co-operation and development: Principles and Guidelines for Access to Research Data from Public Funding, 2007, URL: <http://www.oecd.org/dataoecd/9/61/38500813.pdf>, , abgerufen am 20.02.2011
- SURF: Enhanced Publication URL: http://www.surffoundation.nl/nl/publicaties/Documents/SHAREflyer_verrijkte_publicatie_pdfversie_def_ENG.pdf, 2010, abgerufen am 22.02.2011
- Pfeiffenberger, H. ; Carlson, D.: Earth System Science Data (ESSD) - A Peer Reviewed Journal for Publication of Data. In: D-Lib Magazine Bd., 17 2011, Nr. 1/2
- PLoS ONE: PLoS ONE Editorial and Publishing Policies. Sharing of Materials, Methods, and Data. URL: <http://www.plosone.org/static/policies.action#sharing>, 2010, abgerufen am 22.02.2011
- Pöschl, U.: Interactive Open Access Publishing and Peer Review: The Effectiveness and Perspectives of Transparency and Self-Regulation in Scientific Communication and Evaluation. In: *Liber Quarterly* 19, 2010, Nr. ¾
- Zacharias, M. C.: Scientists Seeking NSF Funding Will Soon Be Required to Submit Data Management Plans, 2010. URL: http://www.nsf.gov/news/news_summ.jsp?cntn_id=116928, abgerufen am 18.02.2011

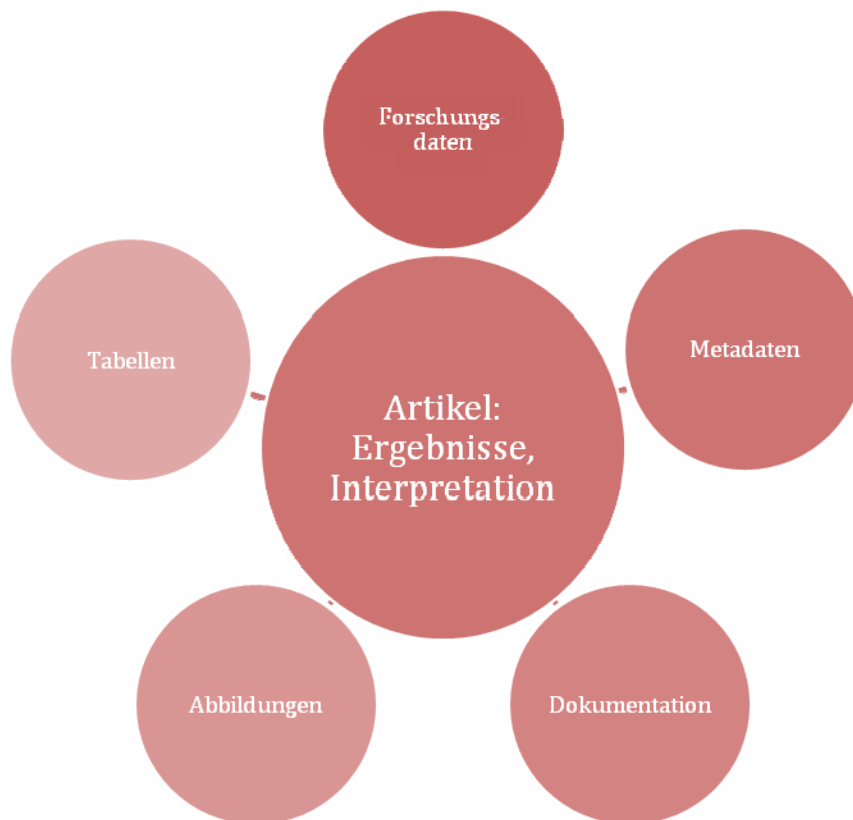
Abbildungen

Abbildung 1



Die Publikation von Forschungsdaten mit einer erweiterten (begutachteten) Dokumentation und möglicherweise weiteren Materialien, welche die Nutzung ermöglichen. Weitere Services können an die Publikation gekoppelt werden. Zusätzliche Analysen und (mehrere) interpretative Artikel können auf den Forschungsdaten aufbauen.

Abbildung 2



Die Publikation von Forschungsdaten als Anhang einer interpretativen Publikation. Die Forschungsergebnisse werden beispielsweise in einem Journal publiziert und die zugrunde liegenden Forschungsdaten zusammen mit dem Artikel bereitgestellt. Im Anhang zu dem interpretativen Artikel finden sich traditionell auch oftmals Abbildungen, Tabellen und detailliertere Angaben zur Methodik.