

Bertelmann, R., Pfeiffenberger, H. (2015): Forschungsdaten und Bibliotheken.  
- In: Griebel, R., Schäffler, H., Söllner, K. (Eds.), *Praxishandbuch Bibliotheksmanagement*, Berlin : De Gruyter, p. 639-651.

## Forschungsdaten und Bibliotheken

### 1 Einführung

Forschungsdaten sind ein traditioneller Teil wissenschaftlichen Arbeitens, der allerdings in der Vergangenheit eher im Schatten anderer wissenschaftlicher Artefakte stand. Trotz oder gerade wegen der quantitativen Zunahme der Forschungsdaten vertiefte sich dieser „Schatten“ im Bereich der kommunizierten wissenschaftlichen Erkenntnis, nachdem der Abdruck relevanter Forschungsdaten im Kontext von Zeitschriftenartikeln aus Platz- und Kostengründen sukzessive abgeschafft wurde. Zwar gab es fachspezifische Datensammlungen, die allerdings nur eine begrenzte Reichweite hatten. Exemplarisch sind hier die Weltdatenzentren, die nach dem Internationalen Geophysikalischen Jahr ab 1957 entstanden.<sup>1</sup> Wenn Forschungsdaten außerhalb fachspezifischer Zirkel überhaupt eine Rolle spielten, dann bestimmten etwa in den neunziger Jahren des 20. Jahrhunderts eher Fragen der Nachvollziehbarkeit mit dem Fokus der Beweissicherung für die Untersuchung wissenschaftlichen Fehlverhaltens die Diskussion. Dies lässt sich seit 1998 ablesen an der Empfehlung 7 der von der Deutschen Forschungsgemeinschaft (DFG) verabschiedeten „Regeln guter wissenschaftlicher Praxis“: „Primärdaten als Grundlagen für Veröffentlichungen sollen auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden, für zehn Jahre aufbewahrt werden.“<sup>2</sup>

Die Berliner Erklärung zum offenen Zugang zu wissenschaftlichem Wissen hat dann bereits 2003, neben ihren Verdiensten für die Entwicklung des Open Access von Textpublikationen, den Begriff des wissenschaftliches Wissens weiter gefasst und Forschungsdaten gleichberechtigt neben der textuellen Publikation in den Blick genommen: „Open access contributions include original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material.“<sup>3</sup>

---

<sup>1</sup> Heute aufgegangen im World Data System (WDS), <http://www.icsu.org/what-we-do/interdisciplinary-bodies/wds/?icsudocid=about>.

<sup>2</sup> DFG 2013: 21. Die erste Auflage des Dokuments von 1998 liegt seit 2013 in einer ergänzten und aktualisierten Version vor.

<sup>3</sup> [http://openaccess.mpg.de/3515/Berliner\\_Erklaerung](http://openaccess.mpg.de/3515/Berliner_Erklaerung).

Im Kontext der Diskussion um Open Access wurde damit die Zugänglichkeit von Forschungsdaten als Thema gesetzt, dem davor lange Zeit wenig Aufmerksamkeit geschenkt worden war. Auch dem noch darüber hinausgehenden Konzept der Publikation von Forschungsdaten wurde damit Legitimation zuteil. Dies fügte sich ein in eine stetig anwachsende, bis heute anhaltende und noch längst nicht abgeschlossene Diskussion zur Neuausrichtung des Verständnisses der Rolle und der Möglichkeiten wissenschaftlicher Daten.

Ein Paradigmenwechsel beim wissenschaftlichen Arbeiten wird beschworen.<sup>4</sup> Forschungsdaten, in der Vergangenheit ein eher „graues“ Material, rücken ins Zentrum wissenschaftlichen Arbeitens. So spricht die National Science Foundation in den USA von „Produkten“ anstelle von Publikationen als zu erwartender Ergebnisse ihrer Förderung und wertet damit explizit Forschungsdaten auf gleicher Ebene wie klassische Publikationen. Europäische Forschungsorganisationen und Forschungsförderer verdeutlichen 2009 in einer Roadmap, dass es in der Forschung vor allem um den offenen Zugang zu qualitätsgesicherten Daten geht.<sup>5</sup> Schlagworte vom neuen Gold machen die Runde.<sup>6</sup> *Share, Re-use*, die Verknüpfung verschiedenster Daten- und Textkörper und avancierte Auswertungsmöglichkeiten (*Datamining*; virtuelle Forschungsumgebungen) markieren die Leitlinien der künftigen Entwicklung.

Dies birgt zahlreiche Herausforderungen nicht nur technischer Art. Rechtliche Fragen von Datenschutz bis zum Urheberrecht stellen sich hier auf der einen Seite. Andererseits generieren solche Anforderungen grundlegenden Veränderungsbedarf in sozio-kulturellen Fragen des wissenschaftlichen Arbeitens.

Die Allianz der Deutschen Wissenschaftsorganisationen hat 2010 als Reaktion auf diese Entwicklung entsprechende Grundsätze zum Umgang mit Forschungsdaten aufgestellt<sup>7</sup>. Folgende Themenfelder wurden als Zukunftsaufgaben gesetzt:

- Sicherung und Zugänglichkeit,
- Berücksichtigung der Unterschiede der wissenschaftlichen Disziplinen,
- Wissenschaftliche Anerkennung (vgl. Kapitel 4),
- Lehre und Qualifizierung,
- Verwendung von Standards,
- Entwicklung von Infrastrukturen.

---

<sup>4</sup> Hey 2009.

<sup>5</sup> EUROHORCS 2009.

<sup>6</sup> Kroes 2011.

<sup>7</sup> Allianz 2011.

Die Royal Society beschäftigt sich im Bericht „Science as an open enterprise“<sup>8</sup> mit den Chancen, aber auch mit Haken und Ösen dieser Entwicklung. Bemerkenswert ist die Setzung des Begriffs *Intelligent Openness*, der als Lösungsweg für einige der anstehenden Fragen geprägt wird. Die differenzierte Behandlung der Offenheit nimmt nicht nur selbstverständliche Grenzen der Offenheit – etwa bei medizinischen Daten – zur Kenntnis, sondern fordert mit den Stichworten *accessible, assessable, intelligible, useable* auch eine Mindestqualität der (offenen) Daten.

## 2 Domänen und Stakeholder

Der Umgang mit Forschungsdaten wird oft als Kontinuum dargestellt.<sup>9</sup> Forschungsdaten sind Rohstoff, Produkt und Ergebnis im wissenschaftlichen Arbeiten und erfordern je nach Standpunkt und Anforderungen unterschiedliche Herangehensweisen. Die folgende Grafik untergliedert entsprechend die verschiedenen Domänen beim Umgang mit diesen Daten.

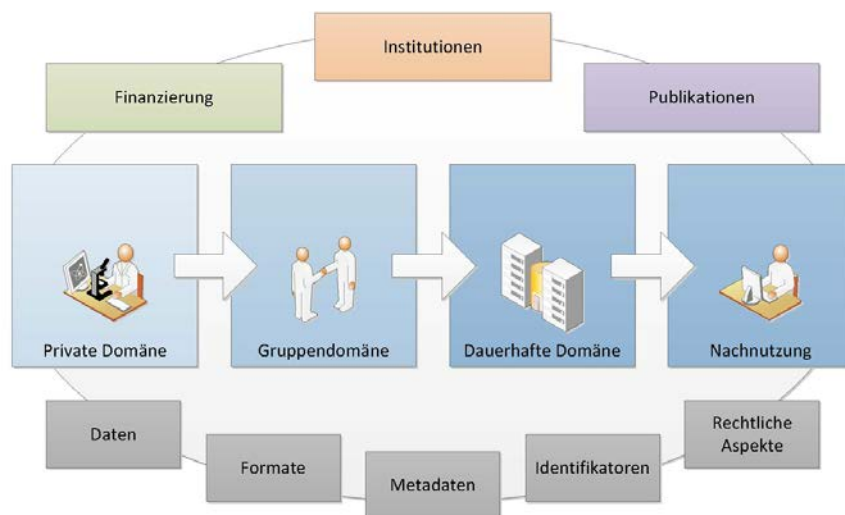


Abbildung 1: Domänen des Datenmanagements<sup>10</sup>

<sup>8</sup> Royal Society 2012.

<sup>9</sup> Treloar 2008: 6.

<sup>10</sup> Abbildung 1 ist entnommen aus: DFG-Projekt RADIESCHEN – Rahmenbedingungen einer disziplinübergreifenden Forschungsdateninfrastruktur (Ed.) 2013: Organisation und Struktur, 61 S. DOI:10.2312/RADIESCHEN\_005. – Vgl. auch: ZfBB 2013: Klar, J., Encke, H.: Forschungsdaten in der Gruppendomäne.

Hier wird klar, dass in den verschiedenen Domänen sehr unterschiedliche Anforderungen an den Umgang mit Forschungsdaten bestehen. Dies bedingt unterschiedliche Akteure (Stakeholder) in der jeweiligen Domäne. Im Diskurs über Forschungsdaten ist eine Verortung des jeweiligen Gesprächsschwerpunkts in diesen Domänen anzuraten, um das jeweilige Aufgabenspektrum spezifizieren und die zugehörigen Akteure identifizieren zu helfen. Die beiden Domänen *Private Research Domain* (Private Domäne) und *Shared Research Domain* (Gruppendomäne) sind beispielsweise stark in die Arbeitsabläufe von Wissenschaftlern eingebunden und werden entsprechend von diesen geprägt. Dagegen wird unter Umständen die Verantwortung für die Daten in den Domänen *permanent* (dauerhaft) und *public* (öffentliche Nachnutzung) auf andere, institutionelle Gruppen übergehen, die an dieser Stelle klassische Dienstleistungen für den Wissenschaftsbetrieb erbringen. Die Verbindung der Domänen und ihre fließenden Übergänge ergeben das Gesamtbild einer anzustrebenden Forschungsdateninfrastruktur, die in Zukunft optimale Bedingungen für den Umgang der Wissenschaft mit Forschungsdaten bieten soll.

Die öffentliche Domäne umfasst natürlich die Forschungsdatenpublikation, hierunter fallen aber auch traditionelle, eventuell seit vielen Jahren bestehende Systeme und Institutionen, die Forschungsdaten sammeln und in sehr verschiedener Art und Weise zu unterschiedlichen Bedingungen zumindest zugänglich machen. Das Portal *re3data.org* ermöglicht mit Hilfe von *Icons* (u.a. zu Open Access, Persistent Identifier, Licences, Certificates and Standards) einen differenzierten Blick auf bereits bestehende Datenrepositorien unter Aspekten wie Zugänglichkeit und Publikation.

Die besprochenen Domänen sind voneinander abhängig, d.h. wenn etwa bereits in der privaten Domäne eine mögliche Publikation mit bedacht wurde, sind Aufwand und Kosten sicherlich geringer.

Nicht alle im Kontinuum erfassten Daten sind unter Umständen zur Publikation gedacht oder geeignet. Die Differenzierung etwa zwischen temporären Arbeitsergebnissen und Daten, die zumindest im Sinn der guten wissenschaftlichen Praxis archiviert werden müssen, ist von Fachgebiet zu Fachgebiet neu auszuloten.

Wenn folgende Betrachtungsweisen des Themas Forschungsdaten als Matrix über Abbildung 1 gelegt werden, entsteht erst das Gesamtbild in seiner ganzen Komplexität, das Datenökosystem:<sup>11</sup>

- Etabliertes Vertrauen (bezogen auf Daten, Systeme, Menschen)
- Daten sind entdeckbar (*discoverable*)
- Daten werden erhalten (*preserved*)
- Daten sind vom Grundsatz her offen und für Menschen und Maschinen zugänglich
- Daten sind nachnutzbar und verständlich aufbereitet

---

<sup>11</sup> Parsons 2013.

- Datensysteme werden effektiv und verteilt organisiert und verwaltet (*governance*)
- Angemessene Wahrnehmung, Wertschätzung und Einbindung des Erstellens, Sammelns und Kuratierens von Daten in Wissenschaftsmetriken.

Sowohl im Begründungskontext des Reports der Royal Society als auch in der Praxis von Erhaltung, Re-Use und Datenpublikation wird darüber hinaus zunehmend das Thema der Qualität diskutiert, was zu einem weiteren Kriterium führt:

- Daten sind nachvollziehbar und damit von belegter Qualität.

### 3 Rollenperspektiven

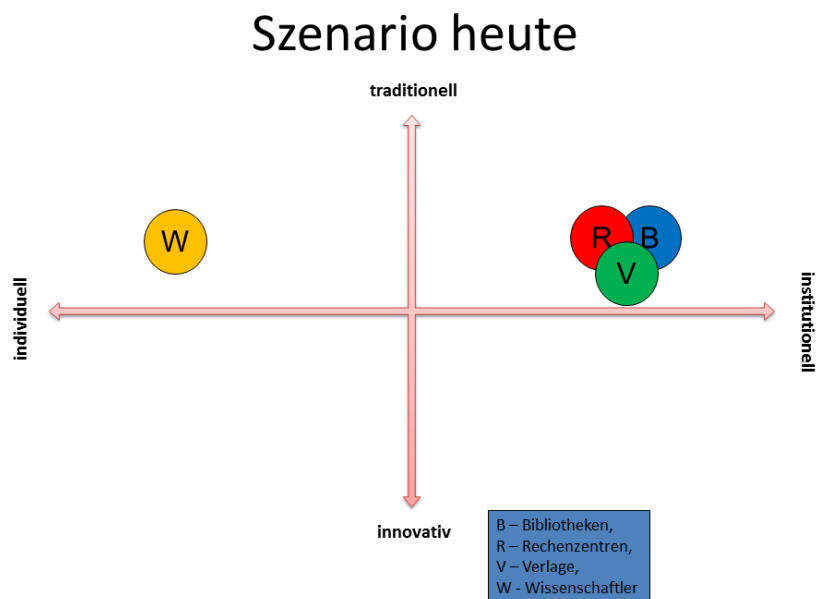
Vor dem Hintergrund dieser vielfältigen Anforderungen auf dem Weg zu einem zukünftigen System der offenen Wissenschaftskommunikation (*Open Science*) verändern sich traditionelle Rollen, und es treten neue Akteure in Erscheinung. So wurde in den letzten Jahren der *Data Scientist* in der Literatur viel diskutiert, und in der Praxis finden sich unterschiedliche Umsetzungen dieses Berufsbilds, insbesondere im Hinblick auf die Zuordnung zur Wissenschaft und/oder zu Informations-Infrastrukturen.

Wissenschaftler, Rechenzentren, Datenzentren und Bibliotheken nehmen bereits Aufgaben in diesem Ökosystem Forschungsdaten wahr. Kommerzielle Anbieter loten die Möglichkeiten von einschlägigen Dienstleistungsangeboten aus. Keiner der Akteure wird im Normalfall das gesamte Spektrum der anfallenden Aufgaben übernehmen können oder wollen. Die Notwendigkeit der arbeitsteiligen Zusammenarbeit liegt daher auf der Hand<sup>12</sup>. Klar ist auch, dass allein zum Zweck, solche Zusammenarbeit von Reibungsverlusten zu befreien, möglichst gut strukturierte und standardisierte Datenmanagementpläne, wie sie Forschungsförderer neuerdings fordern<sup>13</sup>, unabdingbar sind. Momentan lässt sich aber noch nicht absehen, wie in einigen Jahren die Gewichte zwischen den Akteuren verteilt sein werden.

---

<sup>12</sup> Wie in den vom Projekt ODE in „Briefing sheets“ herausgearbeitet, <http://www.alliancepermanentaccess.org/index.php/community/current-projects/ode/outputs/>.

<sup>13</sup> Guidelines on Data Management in Horizon 2020, Version 1.0, 11 December 2013, [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf).

Abbildung 2: Szenario heute<sup>14</sup>

Es ist davon auszugehen, dass eine Neudefinition der Rollen nicht in linearer Form voran schreitet, sondern eher disruptiv, d.h. in dynamischen, unvorhergesehenen Sprüngen.

#### 4 Publikation von Forschungsdaten

Die Generierung von Forschungsdaten erfordert in vielen Fällen aufwendige, teure Infrastrukturen und nimmt einen beträchtlichen Teil der Arbeitszeit der beteiligten Wissenschaftler in Anspruch. In der Vergangenheit war auf Datenquellen allenfalls in Danksagungen in Fußnoten oder am Ende eines Artikels pauschal hingewiesen worden. Wenn Daten doch etwa für Zwecke der Forschungsevaluation zur Verfügung gestellt wurden, konnten zumeist weder die bereitstellende Institution und noch weniger der beteiligte Wissenschaftler von einer möglichen Nachnutzung profitieren. Dies liegt nicht nur an mangelnden Metriken zur Evaluierung von Forschung, sondern grundlegend an einer noch nicht durchgängig bestehenden wissenschaftlichen

<sup>14</sup> Abbildung 2 ist entnommen aus: DFG-Projekt RADIESCHEN – Rahmenbedingungen einer disziplinübergreifenden Forschungsdateninfrastruktur (Ed.) 2013: Synthese, 32 S. DOI:10.2312/RADIESCHEN\_007.

Wertschätzung und Kultur des Zitierens von publizierten Daten. Vor diesem Hintergrund unterbleibt vielfach nicht nur die offene Verfügbarmachung, sondern auch bereits die aufwendige Aufbereitung und Beschreibung der Daten in einer Form, welche eine Nachnutzung durch Dritte überhaupt erst ermöglicht.

Im Lauf der letzten Jahre hat sich daher aus den vielen Perspektiven als elementare Grundlage der skizzierten goldenen Zukunft der Forschungsdaten die Publikation derselben herauskristallisiert. Publikation meint dabei die Aufbereitung der Forschungsdaten analog zu den Publikationsstandards der wissenschaftlichen Textpublikation. Wichtiger Meilenstein auf diesem Weg war die Eröffnung der Möglichkeit, den in der digitalen Textpublikation etablierten, weltweit genutzten persistenten Identifikator DOI (Digital Object Identifier) auch für Forschungsdaten einzusetzen. Die Idee und praktische Umsetzung dieses nun durch die weltweite Organisation DataCite<sup>15</sup> repräsentierten Wegs war im DFG-geförderten Projekt „Publication and Citation of Primary Scientific Data“ (STD-DOI, 2003–2009)<sup>16</sup> entstanden.

Die Publikation von Forschungsdaten hat dabei zwei Ziele. Durch eine Publikation werden Datenpakete sichtbar und eindeutig referenzierbar zugänglich. Grundvoraussetzung ist dabei die Zuordnung elementarer Metadaten im Stil traditioneller Textpublikationen (Autoren, Titel, Erscheinungsjahr, Quelle). Diese normierte Sichtbarkeit ermöglicht ein klares Referenzieren bei der Nachnutzung, ein klassisches Zitat entsteht (wie es etwa die National Science Foundation NSF für Produkte der Forschung fordert). Die Normierung ermöglicht darüber hinaus, dass eine Vielzahl etablierter fachspezifischer Datensammlungen damit auch interdisziplinär sichtbar und nutzbar wird.

Ein zweites Ziel der Datenpublikation betrifft Form und Qualität des Produkts. Ein simples binäres Objekt – selbst wenn es über einen DOI referenzierbar und zugreifbar gemacht ist – ermöglicht noch keine vertrauensvolle Nutzung, ja nicht einmal eine Überprüfung. Neben Angaben zu Autor, Titel etc. ist als Minimum ein ausreichendes Maß an Metadaten nötig, welche Herkunft der Daten und Methoden ihrer Gewinnung genau nachvollziehbar machen. Vielfach werden diese Metadaten auf – eventuell andernorts – veröffentlichte Dokumente verweisen, die diese Information bereitstellen. Die Metadaten könnten auch Referenzen auf Ursprungsdatensätze bereitstellen, wenn etwa der beschriebene Datensatz eine Aggregation oder statistische Zusammenfassung eines oder mehrerer anderer Datensätze ist.

---

<sup>15</sup> <http://www.datacite.org/>.

<sup>16</sup> Projektpartner: TIB Hannover, WDC-MARE (Bremen/Bremerhaven), WDC Climate (Hamburg), GFZ Potsdam, WDC-RSAT (Oberpfaffenhofen). Vgl. auch Klump 2006.

Erst durch diese Abbildung der wissenschaftlichen Methode können Daten zum integralen Teil der “Records of Science” werden und in Zukunft gleichberechtigt neben Textpublikationen stehen und nachgewiesen werden<sup>17</sup>.

## 5 Typologie der Datenpublikation

### 5.1 Daten zur Textpublikation (Supplement)

Datensupplemente zu Textpublikationen haben eine lange Tradition. Aus Kostengründen verschwanden sie im Lauf der Jahre aus den damals noch ausschließlich papiergebundenen Zeitschriften. Trotz der digitalen Möglichkeiten sehen es die meisten Verlage nicht als ihre Aufgabe an, das Format Datensupplement im Rahmen ihrer Produkte zu entwickeln. Zwar wird oft eine Möglichkeit zur Ablage von ergänzendem Material angeboten, auf vielen Verlagsplattformen führen diese aber ein eher stiefmütterliches Dasein. Andererseits nehmen auch Verlage die zunehmende Bedeutung von Forschungsdaten wahr, fordern zugehörige Supplementdaten ein, verweisen zur Publikation aber auch auf von Dritten betriebene Datenrepositorien. In den Policies der Zeitschriften aus der Nature-Familie ist dies so formuliert:

An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. Therefore, a condition of publication in a Nature journal is

- that authors are required to make materials, data and associated protocols promptly available to readers without undue qualifications.
- Supporting data must be made available to editors and peer-reviewers at the time of submission for the purposes of evaluating the manuscript.”<sup>18</sup>

Die Zeitschriftenfamilie des Copernicus-Verlags geht bereits weiter und erwartet<sup>19</sup> „...depositing data that correspond to journal articles in reliable data repositories, assigning digital object identifiers, and properly citing a data set as a proper citation.“

<sup>17</sup> Beispiel Datenpublikation: The global carbon budget 1959–2011 In: Earth System Science Data (2013) 5, S. 165–185. doi:10.5194/essd-5-165-2013), interpretierende Publikation: The challenge to keep global warming below 2 °C. In: Nature Climate Change, 3, 4–6 (2013). doi:10.1038/nclimate1783.

<sup>18</sup> <http://www.nature.com/ngeo/authors/index.html>.

<sup>19</sup> etwa bei der Zeitschrift „Soil“: [http://www.soil-journal.net/general\\_information/data\\_policy.html](http://www.soil-journal.net/general_information/data_policy.html).



## 5.2 Selbständige Datenpublikation analog zu traditionellen Publikationstypen

Bereits existierende Datenrepositorien ertüchtigen ihre Anwendungen, um alle oder einzelne Datenpakete in ihren Systemen publizierbar und zitierbar zu machen. Beim Neuaufbau eines Datenrepositoriums wird dies natürlich Dreh- und Angelpunkt des Konzepts sein. In den letzten Jahren ist eine ganze Reihe von sogenannten Datenzeitschriften entstanden<sup>20</sup>. Dies sind Zeitschriften, deren Fokus auf der Beschreibung selbständiger Datenkonglomerate liegt, die zuvor in einem zuverlässigen Datenrepositorium publiziert worden sind.

Auf eine ähnliche Weise eignen sich Formate aus der klassischen grauen Literatur wie institutionelle Reportreihen<sup>21</sup> für eine solche Art der Publikation. Persistent elektronisch bereitgestellt, können diese auf neue Art wieder Eingang in den Kanon wissenschaftlicher Kommunikation finden.

Entscheidend ist bei alledem, dass die Verbindung der nackten Daten mit ihren Vorläufern, qualitätssichernden Informationen, aber auch mit den Interpretationen (Publikationen), die auf sie aufbauen, hergestellt und erhalten wird.

Bei aller Nachbildung der Mechanismen der Textpublikation, z.B. durch *Peer Review* auf die Daten selbst, darf nicht übersehen werden, dass Daten vielfach einen anderen Lebenslauf haben als Textpublikationen. Während letztere nach der formalen Publikation quasi unveränderlich sind, ist es nicht unüblich, dass gerade besonders nützliche, z.B. hoch aggregierte Datenprodukte, im Verlauf der Zeit vielfach ergänzt und erweitert, gelegentlich auch durch Anwenden neuer Methoden verbessert werden. Diese Versionierung von Datensätzen vergrößert den Bedarf an Nachvollziehbarkeit und erhöht naturgemäß die Komplexität, diese zu realisieren.

---

<sup>20</sup> Earth System Science Data (www.earth-syst-sci-data.net); Scientific Data (<http://www.nature.com/scientificdata/>).

<sup>21</sup> Beispiel: Möller, F., Liebscher, A., Martens, S., Schmidt-Hattenberger, C., Kühn, M. (2012): Yearly operational datasets of the CO<sub>2</sub> storage pilot site Ketzin, Germany, (Scientific Technical Report Data, 12/06), Potsdam, Deutsches GeoForschungsZentrum GFZ, 14 S., 10.2312/GFZ.b103-12066.

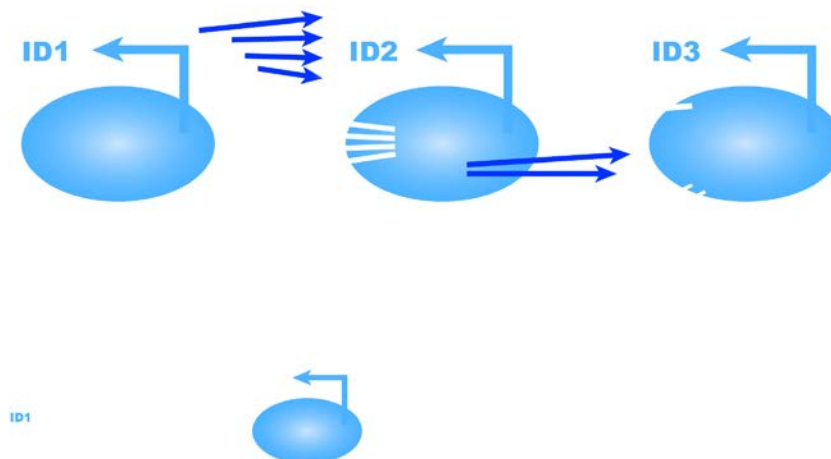


Abbildung 3: Objekte der Datenpublikation

Neben den eigentlichen Objekten, nämlich den Texten und Daten in ihren angeborenen (binären) Formen und Formaten, gibt es hinzugefügte Metadaten aller Art, wobei die Verbindung zwischen den binären Daten, den Metadaten und den verschiedenen Objekten im Publikationskosmos durch persistente Identifier aufrecht erhalten wird. Die Abbildung zeigt nur die Verlinkung – sprich: Zitiermöglichkeiten – im einfachsten der denkbaren Szenarien mit Datenjournal.

## 6 Die Rolle der Bibliotheken

### 6.1 Relevante bibliothekarische Kompetenzen

In den Sozialwissenschaften wie auch im Bereich der MINT-Fächer hatten wissenschaftliche Bibliotheken in der Vergangenheit nur wenige Berührungspunkte mit dem Komplex Forschungsdaten. Dagegen haben geisteswissenschaftliche Bibliotheken seit jeher eine wichtige Rolle bei der Bereitstellung der Forschungsdaten gespielt. Allerdings wurden etwa historische Quellenmaterialien selten unter dem Begriff „Forschungsdaten“ gefasst, was sich mit dem Aufkommen der sogenannten *Digital Humanities* nun verändert hat.

Forschungsdaten sind grundsätzlich fachspezifisch ausgeprägt. Die konkrete Art der Daten, ihre Einbindung in das wissenschaftliche Arbeiten und der entsprechende Umgang mit Forschungsdaten differiert stark von

Fachgebiet zu Fachgebiet. Vor dieser Folie ist ein fachgebundenes Angebot etwa von Universitätsbibliotheken, die naturgemäß die ganze Breite der am Ort vorhandenen Disziplinen bedienen müssen, nur durch eine entsprechende Schwerpunktsetzung realisierbar. Andererseits gibt es eine ganze Reihe von eher generisch zu lösenden Notwendigkeiten, die für alle Fächer angeboten werden können, beispielsweise die organisatorische Umsetzung der DOI-Vergabe. Spezialbibliotheken, normalerweise bereits als Dienstleister für themenzentrierte Institutionen aufgestellt, können fachspezifisch agieren.

Die zukünftige Rolle von Bibliotheken im Umfeld der Forschungsdaten ist geprägt von ihren klassischen Kompetenzen. Für sie als die zentralen Informationsvermittler einer Institution mit Know-how zu Recherche-Instrumenten und breitem Wissen zu Inhalten liegt die Integration entsprechender Aspekte aus der Welt der Forschungsdaten in das eigene Aufgabenportfolio nahe.

Auch Repositorien gehören inzwischen zu den grundlegenden Werkzeugen, die in Bibliotheken eingesetzt werden. Zahlreiche Bibliotheken haben sich, korrespondierend damit, zu Publikationsdienstleistern entwickelt. Umfassende Aufgaben des Publikationsmanagements, d.h. die Dokumentation für interne und externe Zwecke, aber auch die Dissemination der wissenschaftlichen Publikationen einer Institution werden inzwischen von vielen Bibliotheken übernommen. Vielerorts spielen CRIS-Systeme (Current Research Information System) eine immer wichtigere Rolle. Das wissenschaftliche Produkt „Forschungsdaten“ hier zu integrieren, ist naheliegend.

Wenn Bibliotheken darüber hinaus die Bereitstellung virtueller Forschungsumgebungen als Aufgabe für sich identifiziert haben, werden Forschungsdaten naturgemäß eine zentrale Rolle bei den entsprechenden Aktivitäten spielen.

Mit ihren Kontakten zur Wissenschaft und entsprechenden Serviceangeboten sind Bibliotheken also im Wettbewerb mit anderen Dienstleistern in einer Institution an vielen Orten im klaren Vorteil. Dabei empfiehlt sich grundsätzlich die intensive Zusammenarbeit mit anderen Akteuren innerhalb der jeweiligen wissenschaftlichen Institution ebenso wie Kooperation mit externen Datenrepositorien. So errichtet das in den Niederlanden landesweit agierende Repository für Geistes- und Sozialwissenschaftliche Daten (DANS) zur Zeit sogenannte „Front Offices“ an jeder Universität – sind die Bibliotheken nicht der natürliche Ort für solche Aktivitäten?<sup>22</sup>

---

<sup>22</sup> <http://de.slideshare.net/AndreaScharnhorst/fobo-tpdl2013>.

## 6.2 Mögliche Handlungsfelder

Aufbauend auf diesen Kompetenzen und mit Blick auf die oben angesprochenen Domänen lässt sich eine Vielzahl von Handlungsfeldern für Bibliotheken identifizieren. Einige davon sind vielerorts schon heute zu verwirklichen:

- a) Beratung
  - Thematisierung des Komplexes Forschungsdaten in der Institution (Policies)
  - Bibliothek als einer der Ansprechpartner zu Forschungsdaten in der Institution. Dies kann sich auf den Erstkontakt und die Vermittlung eines Überblicks auf die spezifische Forschungsdateninfrastruktur einer Institution beschränken oder aber ausgebaut werden.
  - Beratung und Information, etwa mit Blick auf Datenmanagementpläne und Datenzeitschriften
  
- b) Nachweis
  - Integration von Forschungsdaten in eigene Suchinstrumente wie lokal verfügbare Discovery-Systeme
  - Integration von Forschungsdatenverzeichnissen in lokale Nachschlagestrukturen (*Clearinghouses*)
  - Nachweis von Forschungsdaten in Systemen des Publikationsmanagements (Generierung persönlicher Listen auf Webseiten, Reporting)
  
- c) Publikation
  - Verknüpfung von Publikationsnachweisen mit zugehörigen Datennachweisen
  - Verknüpfung und Bewahrung qualitätssichernder Dokumentation
  - Lizenznehmer für Daten-DOI, Ansprechpartner der Institution für Registrierungsagenturen
  - Einbindung in lokale Strategien der Langzeitarchivierung
  - Datenpublikation, oft in Zusammenarbeit mit wissenschaftlichen Datenakteuren und technischen Partnern unter Berücksichtigung der oben genannten verschiedenen Typen der Datenpublikation

## 7 Ausblick

Die Diskussion eines neuen Umgangs mit Forschungsdaten steht erst am Anfang, obwohl innerhalb weniger Jahre bereits eine erstaunliche Dynamik zutage getreten ist. Der weitere Weg zu einem neu verstandenen Datenmanagement als Dreh- und Angelpunkt von Open Science ist trotzdem noch lang. Verglichen mit dreihundertfünfzig Jahren Tradition bei der

Entwicklung von Standards bei wissenschaftlichen Zeitschriften hat sich im Lauf der letzten zehn Jahre zwar schon erstaunlich viel entwickelt. Die Diskussion vieler Aspekte steht letztlich aber immer noch am Anfang, ist noch gestaltbar und benötigt Gestaltung. Das Ziel muss sein, Forschungsdaten als integralen Teil der *Records of Science* zu etablieren und klassische Publikation und Forschungsdaten enger zu verknüpfen. Soziokulturelle, rechtliche und technische Aspekte des *Re-Use* und damit verbunden extensive Methoden des *Datamining* zeichnen sich als nächste Schritte ab.

Für die Rolle von Bibliotheken in forschenden Institutionen in ihrem lokalen Umfeld wird das beim Thema Forschungsdaten nicht weniger heißen als: *take part or perish*.

## Literatur

- Allianz der Deutschen Wissenschaftsorganisationen (2010): Grundsätze zum Umgang mit Forschungsdaten.  
<http://www.allianzinitiative.de/de/handlungsfelder/forschungsdaten/grundsaeetze>.
- Becker, P.-N., Fürste, F. (2013): Sollen wir Bibliothekare jetzt alle Informatiker werden? *BuB* 65, H. 7–8, S. 512–514.
- Deutsche Forschungsgemeinschaft, Ausschuss für Wissenschaftliche Bibliotheken und Informationssysteme, Unterausschuss für Informationsmanagement (2009): Empfehlungen zur gesicherten Aufbewahrung und Bereitstellung digitaler Forschungsprimärdaten. Bonn.  
[http://www.dfg.de/forschungsforderung/wissenschaftliche\\_infrastruktur/lis/veroeffentlichungen/dokumentationen/download/ua\\_inf\\_empfehlungen\\_200901.pdf](http://www.dfg.de/forschungsforderung/wissenschaftliche_infrastruktur/lis/veroeffentlichungen/dokumentationen/download/ua_inf_empfehlungen_200901.pdf).
- Deutsche Forschungsgemeinschaft, Kommission Selbstkontrolle in der Wissenschaft (2013): Vorschläge zur Sicherung guter wissenschaftlicher Praxis: Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“; Denkschrift = Proposals for safeguarding good scientific practice. Weinheim: Wiley-VCH.  
[http://www.dfg.de/foerderung/grundlagen\\_rahmenbedingungen/gwp/](http://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/gwp/)
- DFG-Projekt RADIESCHEN – Rahmenbedingungen einer disziplinübergreifenden Forschungsdateninfrastruktur (Ed.) 2013: Organisation und Struktur, 61 S. DOI:10.2312/RADIESCHEN\_005.
- DFG-Projekt RADIESCHEN – Rahmenbedingungen einer disziplinübergreifenden Forschungsdateninfrastruktur (Ed.) (2013): Synthese, 32 S. DOI:10.2312/RADIESCHEN\_007.
- EUROHORCS (2009): EUROHORCs and ESF Vision on a Globally Competitive ERA and their Road Map for Actions. 24 S.  
[http://www.eurohorcs.org/SiteCollectionDocuments/ESF\\_Road%20Map\\_long\\_0907.pdf](http://www.eurohorcs.org/SiteCollectionDocuments/ESF_Road%20Map_long_0907.pdf).
- Hey et al., Hrsg. (2009): Forth paradigm: Data-intensive Scientific Discovery. Microsoft Research.  
<http://research.microsoft.com/en-us/collaboration/fourthparadigm/contents.aspx>.
- Klump, J., Bertelmann, R., Brase, J., Diepenbroek, M., Grobe, H., Höck, H., Lautenschlager, M., Schindler, U., Sens, I. u. J. Wächter (2006): Data publication in the open access initiative. – *Data Science Journal*, 5, S. 79–83, DOI: 10.2481/dsj.5.79.

- Klump, J., u. R. Bertelmann (2013): Forschungsdaten. In: Kuhlen, R.; Semar, W., u. D. Strauch (Eds.): Grundlagen der praktischen Information und Dokumentation, De Gruyter 28013, S. 575–583.
- Klump, J. (2009): Digitale Forschungsdaten. In: Neuroth, H.; Oßwald, A.; Scheffel, R.; Strathmann, S., u. M. Jehn (Eds.): nestor Handbuch: eine kleine Enzyklopädie der digitalen Langzeitarchivierung, Hülsbusch, Kap.17: S.104–115. <http://nestor.sub.uni-goettingen.de/handbuch/artikel.php?id=72>.
- Kroes, Nellie (2011): Data is the new gold. (Rede) [http://europa.eu/rapid/press-release\\_SPEECH-11-872\\_en.htm](http://europa.eu/rapid/press-release_SPEECH-11-872_en.htm).
- Lyon, L. (2007): Dealing with Data: Roles, Rights, Responsibilities and Relationships . Report. Consultancy [http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing\\_with\\_data\\_report-final.pdf](http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf).
- National Science Foundation (2007): Cyberinfrastructure. Vision for 21st Century Discovery (NSF 07-28). Arlington. <http://www.nsf.gov/pubs/2007/nsf0728/index.jsp>.
- Parsons, M. A., u. P. A. Fox, (2013): Is Data publication the right metaphor? Data Science Journal, 12. WDS32-WDS46. DOI 10.2481/dsj.WDS-042.
- Pampel, H., Bertelmann, R., u. H.-C. Hobohm (2010): „Data Librarianship“ – Rollen, Aufgaben, Kompetenzen. In: Hohoff, U., u. C. Schmiedeknecht (Hrsg.): Ein neuer Blick auf Bibliotheken. Olms 2010, S. 159–176.
- Pfeiffenberger, H. (2007): Offener Zugang zu wissenschaftlichen Primärdaten. In: Zeitschrift für Bibliothekswesen und Bibliographie, 4/5. S. 207–210. <http://epic.awi.de/Publications/Pfe2007d.pdf>.
- Royal Society (2012): Science as an open enterprise. <http://royalsociety.org/policy/projects/science-public-enterprise/report/>.
- Swan, A.; Brown, S. (2008): Skills, Role & Career Structure of Data Scientists & Curators. Assessment of Current Practice & Future Needs. Report to the JISC. Truro. <http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dataskillscareersfinalreport.pdf>.
- Treloar, A., u. C. Harboe-Ree (2008). "Data management and the curation continuum: how the Monash experience is informing repository relationships". Proceedings of VALA 2008, Melbourne. [http://www.valaconf.org.au/vala2008/papers2008/111\\_Treloar\\_Final.pdf](http://www.valaconf.org.au/vala2008/papers2008/111_Treloar_Final.pdf).
- ZfBB Zeitschrift für Bibliothekswesen und Bibliographie, Schwerpunkt Forschungsdaten 60 (2013), H. 6., S. 295–331.

**Roland Bertelmann** ist Leiter der Bibliothek des Wissenschaftsparks Albert Einstein, eine gemeinsame Bibliothek des Deutschen GeoForschungsZentrums GFZ Helmholtz-Zentrum Potsdam, des Potsdam-Instituts für Klimafolgenforschung, des Alfred-Wegener-Instituts, Helmholtz-Zentrum für Polar- und Meeresforschung Potsdam und des IASS Potsdam, Institute for Advanced Sustainability Studies. Er ist Co-Leiter des Helmholtz Open-Access-Projekts.

**Dr. Hans Pfeiffenberger** ist Leiter der IT-Infrastruktur des Alfred Wegener Instituts, Helmholtz Zentrum für Polar- und Meeresforschung, Bremerhaven, Chief Editor der Zeitschrift „Earth System Science Data“ und Co-Leiter des Helmholtz Open-Access-Projekts.