

Müller, Lars

## Kooperatives Management geisteswissenschaftlicher Forschungsdaten

*formal und inhaltlich überarbeitete Version der Originalveröffentlichung in:*

*formally and content revised edition of the original source in:*

*ABI-Technik 39 (2019) 3, S. 194-201, 10.1515/abitech-2019-3003*



Bitte verwenden Sie in der Quellenangabe folgende URN oder DOI /

Please use the following URN or DOI for reference:

urn:nbn:de:0111-dipfdocs-206207

10.25657/02:20620

<https://nbn-resolving.org/urn:nbn:de:0111-dipfdocs-206207>

<https://doi.org/10.25657/02:20620>

### Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

### Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

### Kontakt / Contact:

DIPF | Leibniz-Institut für  
Bildungsforschung und Bildungsinformation  
Frankfurter Forschungsbibliothek  
publikationen@dipf.de  
www.dipfdocs.de

## Kooperatives Management geisteswissenschaftlicher Forschungsdaten

### Zusammenfassung:

Sobald in geisteswissenschaftlichen Forschungen Daten selbst erzeugt werden, wird Forschungsdatenmanagement erforderlich. Informationsinfrastruktureinrichtungen haben auf ein funktionierendes Datenmanagement mit der Digitalisierung von Kulturgut großen Einfluss, und nicht immer gibt es ein abgestimmtes Rollenverständnis. Die Verständigung über divergierende Sichten bei der Datenerzeugung durch die Domänen Forschung und Infrastruktur ermöglicht erfolgreiches Datenmanagement. Ein aktiver Datenmanagementplan dokumentiert diesen Prozess.

**Schlüsselwörter:** Forschungsdatenmanagement, Digital Humanities, Informationsinfrastrukturen

## Collaborative Research Data Management in the Humanities

### Abstract:

Research Data Management in the humanities is necessary when scholars generate data themselves. In the times of the mass digitization of cultural heritage, information infrastructures can have a big impact on how data is organized. This can cause some problems. Thus, a shared understanding of the roles and aims in data generation will support successful research data management. An active data management plan will document this process.

**Keywords:** Research Data Management, Digital Humanities, information infrastructures

## **1. Einleitung**

Forschungsdatenmanagement in den Geisteswissenschaften war lange ein Nischenthema. Die etablierten Verfahren der Zitation von Quellen, die hermeneutische Deutung von Texten, Bildern oder Objekten erforderte über den klassischen Anmerkungsapparat hinaus keine spezielle Beachtung der „Daten“. Auch die verbreitete Zitation publizierter, digitaler Quellen hat daran nichts geändert.

Erst in jüngerer Zeit verlässt die Diskussion über Forschungsdaten in den Geisteswissenschaften den engeren Zirkel der Digital Humanities und wird zunehmend auch in geisteswissenschaftlichen Fachdisziplinen geführt.<sup>1</sup> Das könnte einem zunehmenden Ineinandergreifen von traditionellen Arbeitsweisen und digital orientiertem Vorgehen geschuldet sein. Die Praxis des Forschungsdatenmanagements, wie sie in den datenintensiven Disziplinen entwickelt wurde und sich inzwischen weitgehend durchgesetzt hat, lässt sich jedoch nicht ohne weiteres auf geisteswissenschaftliche Arbeitsweisen übertragen.

Zur Veranschaulichung sei hier ein idealtypischer Fall skizziert:

Wissenschaftlerinnen und Wissenschaftler, die ein neues Forschungsvorhaben beginnen, kommen auf die Bibliothek zu und wünschen digitalen Zugriff auf die historischen Bestände. Da nicht alles digitalisiert vorliegt, haben die Forschenden idealerweise daran gedacht, Mittel für die Digitalisierung zu beantragen, mit denen die Bibliothek in die Lage versetzt wird, die gewünschten Daten zu erzeugen und auszuliefern. Man trifft sich zum Projektstart und berät über Auswahl, Priorisierung, Vorgehen und Aufgabenverteilung bei der Erstellung der Digitalisate, wird sich einig und beginnt mit der Arbeit.

Im weiteren Verlauf des Projekts zeigt sich dann aber, dass seitens der Infrastruktureinrichtung (Bibliothek oder Archiv) und den Forschenden sehr unterschiedliche Erwartungshaltungen, Sichten und Anforderungen in Bezug auf das digitale Material bestehen: Die Infrastruktureinrichtung legt ihren Fokus auf Vollständigkeit, Nachhaltigkeit, Austauschbarkeit und Nachnutzbarkeit, die Forschung hat einen inhaltlichen, selektiven Fokus und wünscht sich Schnelligkeit, exklusiven Zugriff und Freiheit bei der Auswahl von Software und Formaten.

Hier entsteht ein Spannungsverhältnis, für das wir einen konstruktiven Umgang benötigen. Vor dem Erfahrungshorizont aus der Praxis einer Forschungsbibliothek, der Bibliothek für Bildungsgeschichtliche Forschung (BBF), wird im Folgenden dieses Spannungsfeld abgesteckt und anschließend werden Anregungen zur produktiven Lösung entwickelt.

## **2. Der Forschungsgegenstand wird selbst erzeugt**

Die Notwendigkeit zum Forschungsdatenmanagement in den Geisteswissenschaften entspringt einem Paradigmenwechsel. Vereinfacht gesagt wurden in der analogen Welt einer textorientierten geisteswissenschaftlichen Forschung Quellen kritisch gelesen, ausgewertet und in einem größeren geistigen Zusammenhang gedeutet. Durch die Digitalisierung wurde der zeit- und ortsunabhängige Zugriff auf zahlreiche Quellen verbessert. Bibliotheken und Archive bieten im Internet digitale Surrogate der physischen Originale an. Das Bemühen, mit der digitalen Darstellung die physische nachzuahmen, verschleiert für die Nutzung den Transformationsprozess. Dadurch ist es grundsätzlich möglich, die Quelle so zu verarbeiten, als ob das Original genutzt wurde. Es ist durchaus vorstellbar, dass häufig die Originalvorlage zitiert wird, ohne kenntlich zu machen, dass tatsächlich die digitale

---

<sup>1</sup> Beispielhaft seien hier eine Tagung aus dem Feld der Geschichtswissenschaft und die Thematisierung von Fragen des Forschungsdatenmanagements in den ethnologischen Fächern genannt: Forschungsdaten in der Geschichtswissenschaft, 07.06.2018–08.06.2018 Paderborn, in: H-Soz-Kult, [www.hsozkult.de/conferencereport/id/tagungsberichte-7859](http://www.hsozkult.de/conferencereport/id/tagungsberichte-7859) (15.09.2018); der Fachinformationsdienst (FID) „Sozial- und Kulturanthropologie“ befasst sich mit Forschungsdatenmanagement in den ethnologischen Fächern (<http://www.evifa.de/v2/de/ueber-evifa/forschungsdatenmanagement>).

Reproduktion verwendet wurde. Das ist beim Zugriff auf zuverlässige, z. B. DINI-zertifizierte<sup>2</sup> Repositorien auch weitgehend unproblematisch. Bis zu diesem Punkt ist das traditionelle Arbeiten mit einem Anmerkungsapparat, der verwendete Quellen und Sekundärliteratur auflistet, zweckmäßig und ausreichend. Ein darüber hinausgehendes Management von Forschungsdaten ist nicht erforderlich.

Das verändert sich aber grundlegend in dem Moment, in dem für eine geisteswissenschaftliche Forschung die Daten selbst erzeugt werden. Und genau dies geschieht mit zunehmender digitaler Durchdringung geisteswissenschaftlicher Forschung immer häufiger, d. h., eigentlicher Gegenstand der wissenschaftlichen Betrachtung ist nicht mehr die gelesene Quelle an sich, sondern ein Datensatz, der in einer Kette digitaler Transformationen von Quellen für eine bestimmte Forschungsfrage eigens erzeugt wird.<sup>3</sup>

Für diesen Fall sei hier ein weiteres Beispiel angeführt: Handschriftliche Quellen werden als digitale Faksimiles nachgebildet. Von Forschenden werden sie transkribiert und in TEI/XML übertragen. Eine bestimmte Auswahl der transkribierten Quellen wird aus dem Repozitorium heruntergeladen und als Datensatz zusammengefasst. Unter bestimmter Gewichtung von Metadaten und Inhalten werden diese Daten schließlich mit computerlinguistischen Methoden ausgewertet und das Ergebnis wird in einer Visualisierung graphisch dargestellt.

Die Forschungsfrage bestimmt hier schon die Erstellung der Transkription. Die TEI/XML-Dokumente werden für die Anwendung bestimmter computergestützter Verfahren auf eine Fragestellung hin erstellt. Das Ergebnis zeigt die Auswertung dieses selbst erzeugten Datensatzes. Der Bezug zu den zugrunde liegenden Quellen ist zwar vorhanden, aber er ist nur noch mittelbar. Aus Sicht des Forschungsdatenmanagements ist dieser Fall zumindest dann eher unproblematisch, wenn man ihn isoliert betrachtet. Wäre nur dieser eine, klar definierte Datensatz Hauptgegenstand der daraufhin verfassten Forschungspublikation, könnten bewährte Verfahren aus datenorientierten Wissenschaften für das Forschungsdatenmanagement unverändert übernommen werden.<sup>4</sup>

Dennoch tun wir uns schwer damit, eindeutig zu bestimmen, was geisteswissenschaftliche Forschungsdaten als solche sind, und Forschungsdatenmanagement in die Geisteswissenschaften zu übertragen. Es gibt zahlreiche Ansätze, dieses Problem zu bewältigen. In einem wird z. B. die Digitalität und die dauerhafte Aufbewahrung und Nachnutzbarkeit in das Zentrum gerückt,<sup>5</sup> in einem anderen unter dem Begriff „Quelldaten“<sup>6</sup> die digitale Transformation physischer Quellen besonders beachtet.

Einen umfassenden Ansatz zur Definition geisteswissenschaftlicher Forschungsdaten verfolgt Peter Andorfer, in dem er die Datenpyramide, bestehend aus einer breiten Basis Rohdaten, einer Schicht prozessierter Daten in der Mitte und der Publikation an der Spitze, für die Geisteswissenschaften adaptiert. Hierfür greift er auf die Begriffe Quellen (statt Primär- oder Rohdaten) und Arbeitsdaten (anstelle prozessierter Daten) zurück und fügt Informationsinfrastruktureinrichtungen als eine unter der Pyramide liegende Sockelschicht ein. Zwei aus der Pyramide herausweisende Pfeile symbolisieren, dass die Arbeitsdaten in ein Repository

---

<sup>2</sup> <https://dini.de/dienste-projekte/dini-zertifikat/> (12.07.2019).

<sup>3</sup> Vgl. Cremer, Fabian, Lisa Klaffki, Timo Steyer. „Der Chimäre auf der Spur: Forschungsdaten in den Geisteswissenschaften.“ *o-bib. Das offene Bibliotheksjournal* 5,2 (2018): 144. doi:10.5282/o-bib/2018H2S142-162 (19.11.2018).

<sup>4</sup> Vgl. zum Data Curation Lifecycle: Rümpel, Stefanie. „Der Lebenszyklus von Forschungsdaten.“ In *Handbuch Forschungsdatenmanagement*. Hrsg. von Stephan Büttner, Hans-Christoph Hobohm, Lars Müller, Bad Honnef: Bock + Herchen, 2011, 25–34. <https://nbn-resolving.org/urn:nbn:de:kobv:525-opus-2268> (27.06.2019).

<sup>5</sup> Puhl, Johanna, Peter Andorfer, Mareike Höckendorff u. a. *Diskussion und Definition eines Research Data LifeCycle für die digitalen Geisteswissenschaften*. DARIAH-DE Working Papers 11. Göttingen: Niedersächsische Staats- und Universitätsbibliothek Göttingen, 2015, 14.

<sup>6</sup> Degkwitz, Andreas. „Die Tradition des Textes und die Herausforderung der Daten.“ *Bibliothek Forschung und Praxis* 41,2 (2017): 228. doi:10.1515/bfp-2017-0022.

übertragen werden und die Publikation selbst wieder in die Bibliothek gelangt.<sup>7</sup> Dieses Forschung und Infrastruktur integrierende Verständnis von Forschungsdaten ist grundlegend für geisteswissenschaftliches Forschungsdatenmanagement.

Viel wichtiger als die Frage, was die Daten *sind*, könnte jedoch für geisteswissenschaftliches Forschungsdatenmanagement die Frage sein, woher sie kommen und was mit ihnen *getan* wird. Christine L. Borgman verwies bereits vor einigen Jahren auf zwei wichtige Ansätze, um das Wesen von Forschungsdaten in den Geisteswissenschaften zu verstehen, die an dieser Stelle weiterhelfen:

- 1) Daten sind als formalisierte, zur Interpretation geeignete Darstellung von Information zu betrachten und
- 2) aus ihrem Entstehungszusammenhang heraus zu verstehen.<sup>8</sup>

Der erste Punkt betrifft ganz wesentlich die Modellierung der Daten. „Anders als eine Kopie hat es [das Modell, L.M.] nicht alle Merkmale des Originals, sondern nur ausgewählte. Die Auswahl beruht auf Annahmen des Modellierers, welche Merkmale für die intendierte Verwendung des Modells relevant sind.“<sup>9</sup> Erst die zweckbestimmte Modellierung macht die Daten für die Forschung nutzbar. Selbstverständlich trifft dies auch auf geisteswissenschaftliche Datenmodelle zu. Wenn als Grundlage für eine Forschung Metadaten eines bestimmten Bestandes dienen sollen, hängt von der Modellierung der Daten grundlegend ab, welche Fragen bearbeitet werden können. Ein Datensatz der eine automatische Titelauswertung unterstützt, muss anders gestaltet sein, als einer, auf dessen Basis das quantitative Aufkommen in der Zeit und die geographische Verteilung visualisiert werden sollen.

Mit dem zweiten Punkt stellt Borgman heraus, dass Daten – unabhängig davon, ob sie als Zahlenreihen, numerische Tabellen, Zeichenketten oder strukturierte Texte vorliegen – eine spezifische Herkunft und einen spezifischen Entstehungszusammenhang haben. Borgman nennt hier *Observational Data*, *Computational Data*, *Experimental Data* und *Records*.<sup>10</sup> Für die Geisteswissenschaften sind die *Records*, also prozessproduzierte Daten von besonderer Bedeutung, wenn auch keineswegs die einzige Quellenart.

Um Herkunft und Modellierung der ausgewerteten Daten nachvollziehbar und überprüfbar zu machen, ist bei den selbst erzeugten Forschungsdaten ein systematisches Forschungsdatenmanagement erforderlich. Dies gilt in datenorientierten Disziplinen seit langem und es gibt hier eine gut entwickelte Wissenschaftskultur im Umgang mit Forschungsdaten. In traditionell hermeneutisch arbeitenden Wissenschaften kommt erschwerend hinzu, dass – selbst wenn im Fokus ein bestimmter Datensatz steht – dieser nicht hinreichend für die Forschung ist, sondern in einen größeren Deutungszusammenhang eingefügt werden muss, der sich auf weitere Quellen stützt, darunter potentiell alle Arten physischer wie digitalisierter Objekte der kulturellen Überlieferung.

### 3. Die Infrastruktureinrichtung ist an der Erzeugung der Forschungsdaten unmittelbar beteiligt

Geisteswissenschaften stützen sich stark auf die Bestände von Bibliotheken und Archiven, auch auf die digitalen Bestände dieser Institutionen. Mit der systematischen Kulturgutdigitalisierung sind aber nicht nur die Forschung, sondern auch Bibliotheken und Archive unmittelbar in die Erzeugung von Forschungsdaten involviert. Forschende, die mit digitalisierten Quellen arbeiten, sind dadurch nicht uneingeschränkt souverän bei der Erzeugung ihrer Daten. Es sind die Infrastruktureinrichtungen,

<sup>7</sup> Andorfer, Peter. *Forschungsdaten in den (digitalen) Geisteswissenschaften: Versuch einer Konkretisierung*, Dariah-DE Working Papers 14 (2015): 14. <http://webdoc.sub.gwdg.de/pub/mon/dariah-de/dwp-2015-14.pdf> (12.07.2019).

<sup>8</sup> Borgman, Christine L. „The Digital Future is Now: A Call to Action for the Humanities.“ *Digital Humanities Quarterly* 3,4 (2009): Abs. 25 f. <http://digitalhumanities.org/dhq/vol/3/4/000077/000077.html> (12.07.2019).

<sup>9</sup> Jannidis, Fotis. „Grundlagen der Datenmodellierung.“ In *Digital Humanities: Eine Einführung*. Hrsg. von Fotis Jannidis, Hubertus Kohle, Malte Rehbein. Stuttgart: J.B. Metzler, 2017, 99–108, 100.

<sup>10</sup> Borgman 2009: Abs. 26.

die Bibliotheken und Archive, die die Brücke zwischen physischem Original und dessen digitaler Beschreibung (den Metadaten) sowie der digitalen Repräsentation der Vorlage (dem Digitalisat) errichten. Diese Daten haben einen nachhaltigen und wohldefinierten Bezug zu den Originalvorlagen. Da aber, wie oben bereits erwähnt, die Erzeugung von Daten bestimmten Zwecken und den daraus abgeleiteten Modellierungsprinzipien unterliegt, ist es keineswegs unerheblich, ob die Daten aus einer Infrastruktureinrichtung stammen oder von der Forschung selbst erzeugt werden.

Es gibt Datenmodelle, die geschaffen werden, um bestimmte Forschungsfragen zu modellieren und somit bearbeitbar zu machen. Die Modellierung wird hierbei ganz auf die Forschungsfrage zugeschnitten sein (*research-driven*), und es wäre Zufall, wenn sie für andere direkt nutzbar wäre. Den zweiten Typus bilden dagegen Modellierungen, die extra dazu geschaffen werden, um digitale Ressourcen nachhaltig erschließen zu können. Sie werden häufig von Bibliotheken oder Archiven oder auch von Kooperationen zwischen Fachwissenschaftlern erstellt (*curation-driven*).<sup>11</sup>

Die systematische Digitalisierung von Kulturgut ist *curation-driven*. Die Beschreibung der Objekte mit Metadaten (Katalogisierung und archivarische Erschließung) dient u. a. der Identifizierung bestimmter Dokumente, der Übersicht von Beständen der Recherchierbarkeit etc., die Digitalisierung physischer Vorlagen – also das Abfotografieren und Erzeugen einer digitalen Repräsentation, die das Original möglichst exakt virtuell nachbildet, dient der Zugänglichmachung und der Bestandserhaltung. Die Forschung stützt sich – gezwungener Maßen – auf diese *curation-driven* erzeugten Daten und hängt von ihnen ab, kann sie aber nicht unmittelbar verwenden, sondern „wir [haben] bei der Herstellung von Forschungsdaten stets einen Modellierungsschritt, in dem Gegenstände geisteswissenschaftlicher Forschung in einem Transformationsszenario nach Maßgabe einer geisteswissenschaftlichen Fragestellung in Forschungsdaten umgewandelt werden.“<sup>12</sup>

Forschungsdaten entstehen also erst durch die modellierende Transformation. Das Forschungsdatenmanagement wird durch die Tatsache, dass die Infrastruktureinrichtungen mit der Bereitstellung von Digitalisaten dem Forschungsprozess vorgreifen und direkten Einfluss darauf haben, welche Transformationen zu welchen Forschungsfragen sie ermöglichen, komplexer.

Der Domänenwechsel von der Infrastruktureinrichtung zur Forschung ist in diesem Prozess ein sensibler Punkt, der bislang noch nicht genügend ausgeleuchtet wurde. Vor kurzem verwies Jens Klump auf das Domänenmodell zur Verantwortung im Forschungsdatenmanagement von Treloar und Harboe-Ree. Es beschreibt, wie Forschungsdaten in einem gleichmäßigen Prozess aus der privaten Domäne der Forschung in die öffentliche Domäne der Forschungsdatenrepositorien gelangen. Dabei werden sie mit Metadaten angereichert und einer wachsenden Gruppe von Forscherinnen und Forschern zugänglich gemacht, bis sie schließlich in einem von einer Infrastruktureinrichtung betriebenen Repository publiziert sind.<sup>13</sup> Dieses Modell lässt sich nicht direkt auf geisteswissenschaftliche Forschung übertragen, in der digitalisierte Quellen aus Bibliotheken oder Archiven verwendet werden. Hier beginnt der Prozess in der öffentlichen Domäne, dem Repository, und geht dann erst in die private Domäne der Forschung über.

---

<sup>11</sup> Hervorhebungen im Original. Jannidis 2017: 102.

<sup>12</sup> Stäcker, Thomas. „Noch einmal: Was sind geisteswissenschaftliche Forschungsdaten?“ DHd-Blog. 2015, <http://dhd-blog.org/?p=5995> (07.12.2015).

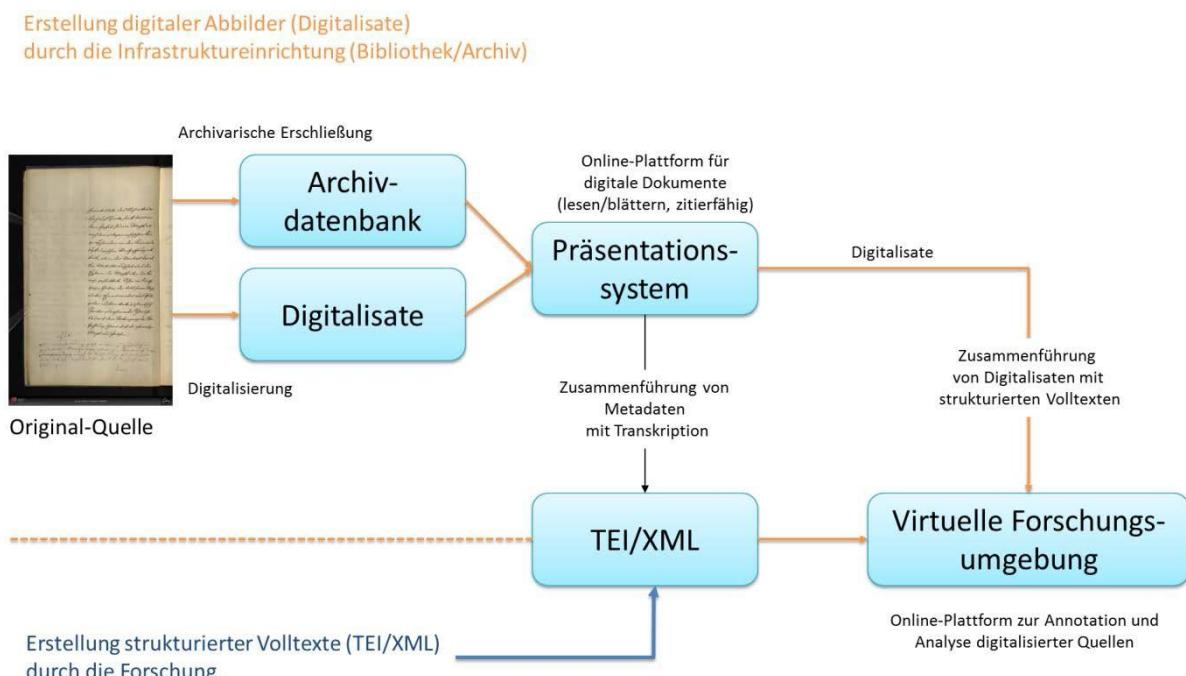
<sup>13</sup> Vgl. Klump, Jens. „Wer ist verantwortlich für das Management von Forschungsdaten, die Forschenden oder die Repositorien?“ *ABI Technik* 38,4 (2018), doi:10.1515/abitech-2018-4015.



**Abb. 1:** Domänen im Verhältnis zur Datenpyramide

In dem hier aufgegriffenen Bild der Datenpyramide (Abbildung 1) steht die weiße Fläche für die digitalen Forschungsdaten eines Projekts. Sie soll verdeutlichen, dass die Daten im Laufe ihres Lebenszyklus und einer bestimmten Forschungsfrage die Domänen wechseln. Die Verantwortung für die Daten „wandert“ zwischen Forschung und Infrastruktureinrichtung – möglicherweise sogar mehrfach – hin und her, und diese führen entweder research-driven oder curation-driven unterschiedliche Transformationen an ihnen aus und stellen sie dann wieder der jeweils anderen Domäne zur weiteren Verwendung zur Verfügung.

Bei der Erzeugung von Forschungsdaten können sich die Daten aber auch zugleich in beiden Domänen befinden, was in der Praxis vermutlich gar nicht so selten ist. Als Fallbeispiel hierfür sei der Prozess zur Datenerzeugung gezeigt, wie er als Workflowmodell in einem bildungshistorischen Projekt entwickelt wurde (Abbildung 2).



**Abb. 2:** Workflow zur Transformation historischer Quellen in Forschungsdaten<sup>14</sup>

<sup>14</sup> Die Grafik basiert auf dem Posterbeitrag: Schindler, Christoph, Julian Hocker, Lars Müller u. a. „Grasping the Materializations of Practices in Digital Humanities, a Semantic Research Environment for Analyzing Exam Grading Practices in German High Schools.“ In *Everything changes, everything stays the same? Understanding*

In einem abgestimmten Vorgehen werden die Daten erzeugt, wobei sich Infrastruktureinrichtung und Forschung mit ihren jeweiligen Expertisen beteiligen. Die Infrastrukturseite übernimmt die Imagedigitalisierung, die Verknüpfung mit den Metadaten (hier die archivarischen Erschließungsdaten aus der Archivdatenbank) und stellt sie als Faksimilepublikation bereit. Die Transkription und Auszeichnung der Quellen in TEI/XML liegt in der Domäne Forschung, da für die Entzifferung der Handschriften und Identifizierung relevanter Entitäten besondere Expertise und erhöhter Aufwand erforderlich sind. Der Aufwand rechtfertigt sich dadurch, dass die Grundlage für die Auswertbarkeit der erzeugten Volltexte gelegt wird. Erst durch die Verknüpfung über einen persistenten Identifier mit den im Präsentationssystem bereitgestellten digitalen Abbildern der Quellen werden die erzeugten Volltexte auch zitierfähig, also in einer geschlossenen Nachweiskette mit dem im Archiv befindlichen Original verbunden. Für die eigentliche Forschungsarbeit können dann die Ergebnisse aus beiden Prozessen in einer Arbeitsumgebung, der virtuellen Forschungsumgebung, zusammengeführt werden.

Dieselben Quellen, gleichzeitig im Blick beider Domänen, haben in der jeweiligen Perspektive einen unterschiedlichen Charakter. Am deutlichsten wird das bei der Betrachtung der Metadaten. In der Archivdatenbank dienen sie der klassischen archivarischen Erschließung. Im Workflow sind sie notwendige Grundlage für Auswahl und Bereitstellung der interessierenden Quellen. Im Zuge der Digitalisierung sind sie unverzichtbar, um das digitalisierte Image mit der Transkription zu verknüpfen. Für die Forschung können die Metadaten aber, sobald sie isoliert für das gesamte Korpus vorliegen, bereits den Charakter eines Forschungsdatensatzes haben. Sie fließen in das Volltextkorpus ein und können in die Analyse (wie z. B. die räumliche und zeitliche Verortung des Korpus) direkt einbezogen werden.

#### 4. Divergierende Sichten und Anforderungen

Die Tatsache, dass die Erzeugung von Daten unter Beteiligung von Infrastruktureinrichtung und Forschung stattfindet, kann zu Missverständnissen, Informationsverlusten und Mehrarbeit auf beiden Seiten führen. „Ein digitales Objekt kann auf drei Ebenen beschrieben werden, als *physisches Objekt*, als *logisches Objekt* und schließlich als *konzeptuelles Objekt*.“<sup>15</sup> Insbesondere auf der Ebene des konzeptuellen Objekts, z. B. des digitalisierten Buchs, kann es bezogen auf das Verständnis digitaler Dokumente zu großen Abweichungen zwischen Infrastruktureinrichtung und Forschung kommen. Für ein funktionierendes Forschungsdatenmanagement ist es wichtig, sich zu vergegenwärtigen, dass die digitalen Objekte dort, wo sie auf physischer und logischer Ebene identisch sind, also in dem Datenspeicher und der Datei, eben nicht automatisch auch auf konzeptioneller Ebene für alle Beteiligten in ihren Rollen identisch sind.

Keine Seite kann bei der Datenmodellierung und -bearbeitung uneingeschränkt die Sicht der anderen übernehmen, weil sie dann ihre jeweiligen Ziele nicht erreichen wird. Wenn die Forschung sich an den Zielen der Infrastruktureinrichtung ausrichtet, besteht für sie die Gefahr, dass der Aufbau der Forschungsinfrastruktur bis zum Projektende dauert, ohne dass die Forschung stattfinden konnte. Übernahme die Infrastruktureinrichtung ganz die Sicht der Forschung, müsste sie den Anspruch auf Erhalt und Weiterentwicklung einer Informationsinfrastruktur aufgeben und arbeitete nicht nachhaltig. Die Tabelle 1 zeigt zugespielt die sich gegenüberstehenden Interessenlagen der beiden Domänen.

---

*Information Spaces: Proceedings of the 15th International Symposium on Information Science (ISI 2017); Berlin, Germany, 13th--15th March 2017.* Hrsg. von Maria Gäde, Violeta Trkulja, Vivien Petras. Schriften zur Informationswissenschaft 70, 365–367.

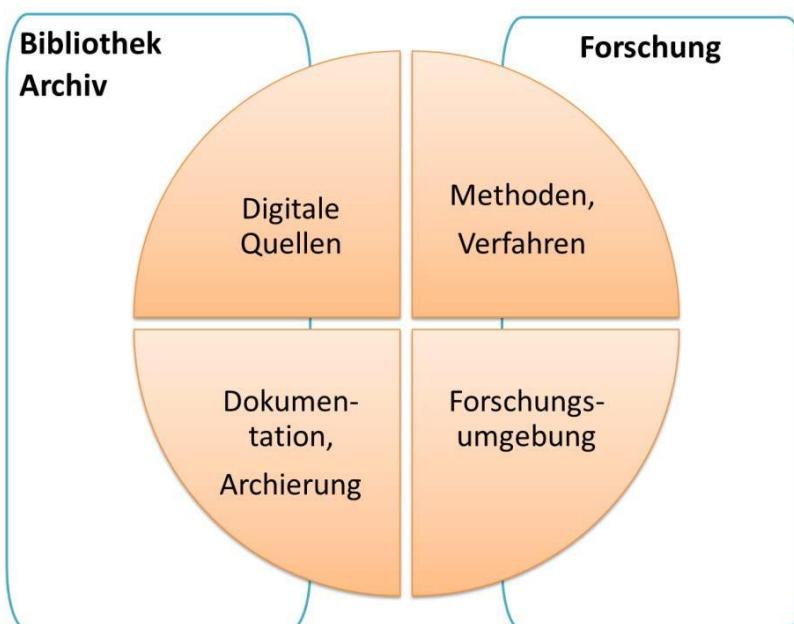
<sup>15</sup> Funk, Stefan E. „Digitale Objekte und Formate.“ In *Nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung*. Hrsg. von Heike Neuroth et al., Version 2.3. Göttingen: Niedersächsische Staats- und Univ.-Bibliothek u. a., 2010, 7:3–7:8.

<http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-20100617136> (12.07.2019).

Perspektive der Forschung	Perspektive von Archiv/Bibliothek
Fokus auf das Forschungsziel	Fokus auf Einbettung in eine globale Informationsinfrastruktur
Zusammenstellung aller thematisch relevanten Quellen	Fokus auf den eigenen Bestand
Sofortiger Zugriff auf die Quellen	Bereitstellung eines sorgfältig aufbereiteten (digitalisierten) und erschlossenen Bestands
Metadaten dienen der inhaltlichen Auswertung	Metadaten dienen der Identifikation und Recherche
Freie Toolauswahl (auch proprietäre Software)	Beschränkung auf langzeitarchivierungsfähige Standards und Austauschformate

**Tab. 1:** Perspektiven im Forschungsdatenmanagement

Betrachtet man die Phasen in einem Forschungsprozess, die auf die Datenbearbeitung zielen, in einem kreisförmigen Modell, wird deutlich, dass das Forschungsdatenmanagement für geisteswissenschaftliche Forschungen ein von Infrastruktur und Forschung gemeinsam und bewusst gelenkter Prozess sein muss. (Abbildung 3) Wichtiger als die jeweilige Gestalt der Daten auf der physischen und logischen Ebene ist dabei die Frage, welche konzeptuellen Modelle der Datenbetrachtung zugrunde liegen und was von wem mit welcher Intention mit den Daten getan wird.



**Abb. 3:** Expertise und Verantwortung im Forschungsdatenmanagement<sup>16</sup>

Das Modell zeigt die beiden Domänen mit ihren jeweiligen datenbezogenen Expertisen. Im Laufe der Forschung „wandert“ die Verantwortung für die Daten und deren Verarbeitung durch die

<sup>16</sup> Die Grafik basiert auf einer Vortragspräsentation: Josefine Wähler und Lars Müller: Digitalisierte Zeitschriften als Quellen der Historischen Bildungsforschung: DGfE-Kongress, Kassel 2016.

verschiedenen Bereiche. Selbst, wenn die Daten unverändert bleiben, können sie im Auge der Betrachter ganz unterschiedlichen Charakter haben. Es kommt darauf an, die Übergänge von einem Bereich in den nächsten und insbesondere von einer Domäne in die andere möglichst gut zu gestalten. Wie gezeigt, greifen die öffentliche Domäne und die private Domäne in einer geisteswissenschaftlichen Forschung schnell ineinander, insbesondere wenn die Forschung die Erzeugung von Digitalisaten zum Zweck einer bestimmten Forschung selbst initiiert, diese aber nach den Gepflogenheiten von Bibliotheken und Archiven von diesen zur öffentlichen Bereitstellung durchgeführt wird.

## 5. Fazit: Forschungsdatenmanagement als Kommunikationsprozess gestalten

In der Literatur werden die zwei Sichten herausgearbeitet, nach denen Bibliotheken entweder als Partner oder Dienstleister der Digital Humanities gesehen werden. Als Partner bringen sie ihre Expertise unmittelbar in die Forschung ein, als Dienstleister stellen sie Quellen, Metadatenformate, Speicherplatz etc. bereit.<sup>17</sup> Für ein funktionierendes Forschungsdatenmanagement ist es dabei notwendig, die momentane Rolle klar vor Augen zu haben und sich über die jeweilige Sichtweise auf die Daten bewusst zu sein. Das geschieht am besten im Dialog. Forschungsdatenmanagement als Kommunikationsprozess heißt:

- Projektbezogener Dialog von Forschung und Archiv/Bibliothek über Daten und Workflows
- Kommunizieren der Anforderungen und Sichten
- Dokumentation der Kommunikationsergebnisse

Die Infrastruktureinrichtung tritt nicht nur als Dienstleister auf, ist aber auch nicht integraler Bestandteil der Forschung. Die Beziehung erhält eher den Charakter einer Datenpartnerschaft, in der beide Domänen in ihren Rollen zusammenwirken. Der Datenmanagementplan (DMP) ist dabei ein zentrales Instrument für „aktives Datenmanagement“.<sup>18</sup> Im Verlauf einer Forschung muss das Forschungsdatenmanagement den tatsächlichen Gegebenheiten wie Schwerpunktverlagerungen oder neu eingebrachte Datentypen angepasst werden. „Dies führt letztendlich hin zu einem aktiven DMP, der über die ganze Laufzeit eines Projektes gepflegt und aktualisiert wird.“<sup>19</sup> Im geisteswissenschaftlichen Projekt dokumentiert er die gemeinsamen und wechselseitigen datenbezogenen Aktivitäten des Forschungsprojekts. Er

- dient der Verständigung über Rollen und Anforderungen,
- ist die Grundlage einer verbindlichen Vereinbarung über den Umgang mit Daten während und nach der Projektlaufzeit und er
- ist der Schlüssel zur Nachnutzung der dafür vorgesehenen Projektdaten.

Forschungsdatenmanagement in einem geisteswissenschaftlichen Projekt kann somit ein strukturierter und bewusst gestalteter Kommunikationsprozess sein. Er beginnt mit der Formulierung einer Policy, die sich aus der institutionellen Policy, den DFG-Empfehlungen zur Sicherung guter wissenschaftlicher Praxis und den Projektzielen ableitet. Iterativ werden im Projektverlauf Standardfragen eines DMP bearbeitet. Welche Datentypen gibt es, welche Rechtsnormen gelten für diese Typen, wer ist verantwortlich für Erhaltung und Nachnutzung etc.? In diese Runden sollten die kuratierenden Infrastruktureinrichtungen und die Institution, an der die Forschung stattfindet, eingebunden sein. Ziel der Iterationen ist es, ein Dokument zu entwickeln, dass zum Projektende die Grundlage für Vereinbarungen zwischen Forschungsprojekt und Infrastruktureinrichtung über den

<sup>17</sup> Vgl. Poole, Alex H., Deborah A. Garwood. „„Natural allies“: Librarians, archivists, and big data in international digital humanities project work.“ *Journal of Documentation* 74,4 (2018): [4 f.]. doi:10.1108/JD-10-2017-0137.

<sup>18</sup> Vgl. Neuroth, Heike, Claudia Engelhardt, Jochen Klar u. a. „Aktives Forschungsdatenmanagement,“ *ABI Technik* 38,1 (2018): 55 ff. doi:10.1515/abitech-2018-0008.

<sup>19</sup> Neuroth 2018: 56 f.

Umgang mit den entstandenen Daten bietet oder selbst diese Vereinbarung darstellt. Zugleich dokumentiert der DMP auch Beschaffung, Verwendung und Verbleib der Forschungsdaten im Projektverlauf. Er stellt somit aus Sicht des Forschungsdatenmanagements einen Forschungsdaten-Metadatensatz dar, der für sich genommen die Datenbasis komplexer Forschungsprojekte transparent abbildet.

Die Anforderungen an das Forschungsdatenmanagement werden am besten gemeinsam erfüllt. Dabei ist es zweckmäßig, ein definiertes Korpus zu haben, das im Zentrum der Forschung steht. Das ermöglicht eine klare Anwendung der Maßnahmen zum Forschungsdatenmanagement auf diesen Bereich. Das Nachhaltigkeitsversprechen der Bibliotheks- und Archivrepositorien kann einen wesentlichen Teil der Dokumentations- und Aufbewahrungspflichten der Forschung abdecken. Dennoch ist nicht jeder Datensatz für die Öffentlichkeit bestimmt. Der Umgang mit den Daten ist schließlich stark von den jeweiligen Datentypen abhängig. In einem DMP, der sich an den üblichen Standards orientiert, werden die Datentypen benannt und Erzeugung, Verwendung, Nachnutzbarkeit etc. sorgfältig dokumentiert. Die Differenzierung von Datentypen ist nötig, um zu erklären, wie mit welchem Datentyp verfahren wurde und wie mit ihm nach Projektende verfahren werden soll. Mögliche Datentypen sind das Kernkorpus, bestehend aus einer einzigen Textgattung, Ergebnisse von Textanalyse-Tools, intellektuelle Annotationen oder Software.

Eine wesentliche Funktion des DMP ist es, die Übergänge zwischen den Domänen und Arbeitsphasen möglichst reibungsarm und ohne Informationsverlust zu gestalten. Hierfür wird festgelegt, wer für welche Daten Verantwortung trägt, wo sie gespeichert werden und zu welchen Bedingungen sie verwendet werden können. In der Regel wird es sich bei der Datengrundlage für eine Forschung nicht um einen einzigen definierten und publizierten Datensatz handeln, der mit einem bestimmten Verfahren verarbeitet wird, woraus ein Ergebnisdatensatz folgt, der unter Einbeziehung des aktuellen publizierten Wissens analysiert wird. Es werden sich im Laufe einer Forschung weitere Bedarfe an zusätzlichen Daten ergeben. In der Praxis muss man davon ausgehen, dass es sich bei der Daten- und Quellengrundlage um heterogenes Material handelt, das sowohl digital als auch physisch vorliegt, und dass Quellen aus verschiedenen Beständen von Archiven oder Bibliotheken stammen und dort ggf. mit unterschiedlichen Parametern digitalisiert wurden.

Somit wird aus dem DMP, der vor Projektbeginn erstellt wurde, eine Forschungsdatendokumentation, deren Aufgabe es ist, alle Informationen aufzunehmen, die notwendig sind, um die Quellengrundlage einer Forschung nachvollziehbar zu machen. Dazu gehört gerade auch die Beschreibung der Verarbeitungsschritte bei der Erzeugung der eigenen Datengrundlage. Denn immer muss der Rückbezug auf die primäre Quelle – unabhängig davon, ob sie ein physisches Objekt oder ein „born-digital“ ist – nachvollziehbar bleiben. Das macht die Verwendung von „Documents as evidence“<sup>20</sup> möglich, die auf eine stabile Beziehung von Informationsinfrastruktur und Forschung gründet.

## Autoreninformationen

Lars Müller

BBF | Bibliothek für Bildungsgeschichtliche Forschung

Abteilung des DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation

Warschauer Straße 34–38

10243 Berlin

[l.mueller@dipf.de](mailto:l.mueller@dipf.de)

[orcid.org/0000-0002-1070-4947](http://orcid.org/0000-0002-1070-4947)

<sup>20</sup> Day, Ronald E. *Indexing it all: The subject in the age of documentation, information, and data*. Cambridge, Mass. u. a.: Mit Press, 2014, 5. Nach: Ficker, Andreas. „Digitale Metaquellen und doppelte Reflexivität.“ H-Soz-Kult, 26.01.2016. <http://www.hsozkult.de/debate/id/diskussionen-2954> (29.01.2016).