

# GOEDOC – Dokumenten- und Publikationsserver der Georg-August-Universität Göttingen

---

---

2015

## Forschungsdaten in den (digitalen) Geisteswissenschaften – Versuch einer Konkretisierung

Peter Andorfer (Herzog August Bibliothek Wolfenbüttel)

DARIAH-DE Working Papers

Nr. 14

Andorfer, Peter: Forschungsdaten in den (digitalen) Geisteswissenschaften : Versuch einer Konkretisierung  
Göttingen : GOEDOC, Dokumenten- und Publikationsserver der Georg-August-Universität, 2015  
(DARIAH-DE working papers 14)

Verfügbar:

PURL: <http://resolver.sub.uni-goettingen.de/purl/?dariah-2015-7>

URN: <http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2015-7-2>

This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)



## Bibliographische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

*Erschienen in der Reihe*  
DARIAH-DE working papers

ISSN: 2198-4670

*Herausgeber der Reihe*  
DARIAH-DE, Niedersächsische Staats- und Universitätsbibliothek

Mirjam Blümm, Thomas Kollatz, Stefan Schmunk und Christof Schöch

---

---

**Abstract:** Die Bezeichnung „Forschungsdaten“ zählt zu den Schlüsselbegriffen der Digitalen Geisteswissenschaften, nicht zuletzt vor den zunehmenden Bemühungen zur Errichtung und Etablierung von Datenzentren, -archiven und Repositorien zur Veröffentlichung und dauerhaften Sicherung von Forschungsdaten. Vorliegendes Papier versucht eine Definition des Begriffes Forschungsdaten. Im Fokus steht dabei das Bemühen, diese Bezeichnung im Vokabular traditionellen geisteswissenschaftlichen Arbeitens zu verorten. Daran anschließend geht es außerdem darum, Forschungsdaten von semantisch ähnlichen Bezeichnungen wie Publikation oder Quellen abzugrenzen.

**Keywords:** Forschungsdaten, Digital Humanities, Geisteswissenschaften  
Research Data, Digital Humanities, Humanities

# Forschungsdaten in den (digitalen) Geisteswissenschaften. Versuch einer Konkretisierung

Peter Andorfer  
(Herzog August Bibliothek Wolfenbüttel)



Peter Andorfer: „Forschungsdaten in den digitalen Geisteswissenschaften. Versuch einer Konkretisierung“. *DARIAH-DE Working Papers* Nr. 14.  
Göttingen: DARIAH-DE, 2015. URN: urn:nbn:de:gbv:7-dariah-2015-7-2.

Dieser Beitrag erscheint unter der  
Lizenz [Creative-Commons Attribution 4.0](https://creativecommons.org/licenses/by/4.0/) (CC-BY).

Die *DARIAH-DE Working Papers* werden von Mirjam Blümm,  
Thomas Kollatz, Stefan Schmunk und Christof Schöch  
herausgegeben.



## Abstract

Die Bezeichnung „Forschungsdaten“ zählt zu den Schlüsselbegriffen der Digitalen Geisteswissenschaften, nicht zuletzt vor den zunehmenden Bemühungen zur Errichtung und Etablierung von Datenzentren, -archiven und Repositorien zur Veröffentlichung und dauerhaften Sicherung von Forschungsdaten. Vorliegendes Papier versucht eine Definition des Begriffes Forschungsdaten. Im Fokus steht dabei das Bemühen, diese Bezeichnung im Vokabular traditionellen geisteswissenschaftlichen Arbeitens zu verorten. Daran anschließend geht es außerdem darum, Forschungsdaten von semantisch ähnlichen Bezeichnungen wie Publikation oder Quellen abzugrenzen.

# Inhaltsverzeichnis

1. Einleitung.....	4
2. Geisteswissenschaften und deren Methoden.....	5
2.1 Geisteswissenschaften der Wissenschaftspolitik.....	5
2.2 Qualitative und quantitative Erhebungen zum Status-Quo der Geisteswissenschaften.....	6
3. Forschungsdaten in den Geisteswissenschaften.....	10
3.1 Spielarten und Definition von Forschungsdaten.....	10
3.2 Beispiele geisteswissenschaftlicher Arbeits- und Forschungsdaten.....	15
3.2.1 – Beispiel A.....	15
3.2.1 – Beispiel B.....	17
3.3 Ein Fazit.....	19
4. Forschungsdaten in den digitalen Geisteswissenschaften.....	20
4.1 Digitale Geisteswissenschaften.....	20
4.2 Repository und Datenzentrum.....	22
5. Fazit.....	24
6. Literaturverzeichnis.....	26

# 1. Einleitung

Die Begriffe Daten, Forschungsdaten, digitale Forschungsdaten bzw. data oder research data haben Hochkonjunktur.<sup>1</sup> Dies dürfte weitgehend unbestritten sein. Erinnerung sei hierfür etwa an den Titel der DHd-Tagung 2015, der bekanntermaßen ja lautete: „Von Daten zu Erkenntnissen“. Weitgehender Konsens dürfte auch über die Gründe dieser Entwicklung herrschen. Häufig genannt werden, erstens die zunehmende Digitalisierung, egal ob darunter jetzt die Digitalisierung von Kulturgütern oder gleich die Digitalisierung der Gesellschaft verstanden wird, zweitens die Forderungen großer Fördergeber nach Sicherung und Veröffentlichung von Forschungsdaten (DFG 2013) und drittens die verschiedenen Initiativen zum Aufbau von Forschungsdatenrepositorien, wobei Letzteres auch eines der Ziele von DARIAH-DE ist.<sup>2</sup>

Die Konjunktur des Themenkomplexes Forschungsdaten lässt sich auch anhand der dazu veröffentlichten Publikation ablesen. So umfasst etwa die für dieses Paper erstellte Literaturliste rund 40 Titel. Dabei gilt es aber zu bedenken, dass vornehmlich Arbeiten darin verzeichnet sind, die das Thema „Forschungsdaten in den Geisteswissenschaften“ untersuchen. Außerdem erhebt diese Literaturliste keinesfalls den Anspruch auf Vollständigkeit. Doch dazu später noch mehr.

Angesichts dieser bereits vorhandenen Publikationen stellt sich natürlich die Frage, ob denn das Thema Forschungsdaten in den Geisteswissenschaften nicht schon erschöpfend behandelt wurde. Mit der Intensivierung der Recherche zeigte sich jedoch, dass man in den Geisteswissenschaften derzeit noch weit von einem einigermaßen allgemein akzeptierten Begriff von Forschungsdaten entfernt ist.

Der Hauptgrund dafür – so eine der Thesen dieser Arbeit – liegt aber weniger an dem Konzept von Forschungsdaten selbst, als vielmehr an einer ehrlichen und kritischen Auseinandersetzung mit dem Wesen der Geisteswissenschaften, vor allem aber mit den Methoden und Arbeitsweisen der Geisteswissenschaften. Ehe daher ein Vorschlag zur Definition von Forschungsdaten in den Geisteswissenschaften vorgestellt wird, sollen zuvor die Charakteristika und die Methoden der Geisteswissenschaften in den Vordergrund gerückt werden. Auf umständliche Exkurse in vergangene Jahrhunderte wird dabei aber ebenso verzichtet, wie auf exzessive Verweise auf Wilhelm Dilthey, Georg Iggers oder Jörn Rüsen. Ausgangspunkt soll vielmehr ein Beitrag aus dem 2012 erschienenen und u.a. von Heike Neuroth herausgegebenen Sammelband *Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme* sein. Und zwar der Beitrag „Geisteswissenschaften“ von Wolfgang Pempe (Pempe 2012).

---

1 Bei dem vorliegenden Paper handelt es sich eine überarbeitete und erweiterte Version des Vortrages „Forschungsdaten in den Geisteswissenschaften. Versuch einer Konkretisierung“, gehalten am 25. Februar im Rahmen der DHd-Tagung 2015 „Von Daten zu Erkenntnissen: Digitale Geisteswissenschaften als Mittler zwischen Information und Interpretation“ in Graz.

2 Vgl. zu weiteren Datenzentren etwa <http://humanities-data-centre.org/> oder <http://dch.phil-fak.uni-koeln.de/startseite.html>.

## 2. Geisteswissenschaften und deren Methoden

### 2.1 Geisteswissenschaften der Wissenschaftspolitik

In den einleitenden Zeilen zu den Ausführungen Wolfgang Pempes finden sich genau jene Phrasen und Schlüsselwörter wieder, die wohl ebenso wenig in einem Projektantrag fehlen dürfen, wie auch in einer Rede zur Lage und zum Stellenwert der Geisteswissenschaften. Demnach sind die Geisteswissenschaften „vielfältig“, sie sind „interdisziplinär“ und außerdem ist „eine Trennung zwischen Geistes- und Naturwissenschaften längst nicht mehr haltbar“. Als konkrete Beispiele für das „Verschwimmen“ der Grenze zwischen den „zwei Kulturen“ nennt Pempe die „Altersbestimmung von Objekten oder die Rekonstruktion mit dem menschlichen Auge nicht (mehr) wahrnehmbarer Schriftzeugnisse“. Im Zuge der Digitalisierung lässt sich außerdem noch eine Zunahme der „ohnehin vorhanden, der Breite des Forschungsbereiches geschuldete[n] Heterogenität“ feststellen (Pempe 2012, S. 137f.). Ein Heterogenität, die in Zukunft „aufgrund der Erweiterung des Spektrums der Methoden als auch der anfallenden Daten noch erhöht wird.“ Nicht vergessen werden darf außerdem der Hinweis, „dass geisteswissenschaftliche Forschungsvorhaben nicht selten als institutionsübergreifende Projekte realisiert [werden]“ (Pempe 2012, S. 140).

Diese Sicht auf die Geisteswissenschaften entspringt aber nicht der Einzelmeinung des Autors. Dieser bezog sich in seinem Beitrag ja einerseits selbst auf andere Texte ähnlichen Inhaltes. So stammt die Aussage vom Verschwimmen der Grenzen zwischen Natur- und Geisteswissenschaften aus einer 2010 veröffentlichten Studie zum Thema „Die internationale Positionierung der Geisteswissenschaften in Deutschland“ (Behrens, Fischer, Minks and Rösler 2010).<sup>3</sup> Andererseits wurden Pempes Ausführungen hinsichtlich Vielfalt, Heterogenität und Interdisziplinarität der Geisteswissenschaften wiederum von späteren Publikationen zitiert (Sahle and Kronenwett 2013; Hügi and Schneider 2013).

Diese vermeintliche Vielfalt und Heterogenität der Geisteswissenschaften – und zwar nicht nur was ihre Fächer, Fragestellungen und Materialien, sondern insbesondere auch was ihre so gerne behauptete Fülle von Methoden und Arbeitsweisen betrifft – mag auf vereinzelte Projekte durchaus zutreffen und soll hier auch gar nicht bestritten werden. Was aber die breite Masse betrifft, die jährlich verfassten Qualifizierungsarbeiten oder die von den nationalen Fördergebern bewilligten Einzelprojekte, so dürften doch Zweifel angebracht sein. Zweifel, aber auch Neugierde, ob es neben diesen Geisteswissenschaften der Antrags- und Wissenschaftspolitik vielleicht noch eine andere Geisteswissenschaft gibt. Ein Geisteswissenschaft, die zwar auf viele Fächer aufgesplittert sein mag, eine Geisteswissenschaft, die zwar die unterschiedlichsten Fragestellungen erforscht, dabei aber womöglich doch einheitlicher vorgeht, mit ähnlicheren Materialien hantiert, weniger vielfältige Methoden verwendet und nicht zuletzt auf eine sehr überschaubare Anzahl von Programmen und Dateiformaten zurückgreift, als dies die wissenschaftspolitische Sicht auf die Geisteswissenschaften vermuten lassen würde.

---

<sup>3</sup> Auf Seite 4, Fußnote 2 ist dabei zu lesen: „Eine scharfe Trennung von Natur- und Geisteswissenschaften erscheint bei manchen neuen Wissenschaftszweigen nicht mehr durchgängig haltbar; die Übergänge – z. B. in der Kognitionswissenschaft verlaufen zum Teil fließend.“

## 2.2 Qualitative und quantitative Erhebungen zum Status-Quo der Geisteswissenschaften

In dem 2013 von Heinrich Geiselberger und Tobias Moorstedt herausgegebenen Sammelband *Big Data* beschreibt Gerhard Lauer, Göttinger Professor für Deutsche Philologie in seinem Beitrag „Die digitale Vermessung der Kultur. Geisteswissenschaften als Digital Humanities“ ein weniger schillerndes, vermutlich aber realistischeres Bild der Geisteswissenschaften (Lauer 2013, S. 116). „Das Rechnen“, so Lauer, „gehörte lange Zeit so gut wie nicht zum Methodeninventar der Geisteswissenschaften, wobei methodische Innovationen in diesem Bereich ohnehin eine eher untergeordnete Rolle spielen.“ Dieses „Methodeninventar der Geisteswissenschaften“ erschöpfe sich, so der Autor weiter, in ihren „historisch-hermeneutischen“ bzw. „qualitativ-interpretierenden Verfahren“. „[U]nd wenn sich doch einmal der Wunsch nach Modernität und Innovation regt, wird dieser durch die theoretischen ‚turns‘ hinreichend bedient.“ Oder: „Eine rechnende Geisteswissenschaft muss vor diesem Hintergrund geradezu als ein Oxymoron erscheinen“ (Lauer 2013, S. 100f.).

Aber wie schlägt sich diese zumindest hinsichtlich der Methodik behauptete Homogenität der Geisteswissenschaften auf die Frage nach dem Wesen der Forschungsdaten nieder? Oder anders formuliert, können die Begriffe Hermeneutik und Daten überhaupt zusammengeführt werden?

Um unter anderem diese Fragen beantworten zu können, werden an der Herzog August Bibliothek eine Reihe qualitativer Interviews mit Gastwissenschaftlerinnen und -wissenschaftlern der Bibliothek geführt. Ein Ziel ist es herauszufinden, mit welchen Programmen die Forscherinnen und Forscher arbeiten und welche Dateiformate sie verwenden. Außerdem bieten diese Interviews eine vortreffliche Gelegenheit, um über das Thema Forschungsdaten zu diskutieren und dafür zu sensibilisieren.

Die Interviews wurden nach einem Leitfaden geführt, die Antworten mit Hilfe eines auf XForms basierenden Web-Formulars erfasst und als XML-Dateien in einer exist-db Datenbank gespeichert. Die Daten können so „on the fly“ mittels xQuery-Skripten analysiert werden.<sup>4</sup> Eine vorläufige Auswertung von 15 Interviews zeigt bisher, dass die Forschenden vornehmlich MS-Word, Literaturverwaltungsprogramme wie Citavi und Zotero, sowie das Dateiverwaltungssystem des jeweiligen Betriebssystems verwenden. Bei den mit Abstand am häufigsten anzutreffenden Dateiformaten handelt es sich um docx, jpeg und pdf. Der Forschungsprozess umfasst außerdem beinahe ausschließlich die Arbeitsschritte, Bibliographieren, Exzerpieren/Annotieren/Notieren und das Schreiben wissenschaftlicher Publikationen in Form von Qualifizierungsarbeiten, Aufsätzen und Monographien (Andorfer 2015, S. 9).

Natürlich können 15 Interviews kein repräsentatives Bild geisteswissenschaftlichen Forschens zeichnen. In weiterer Folge sollen daher verschiedene größer dimensionierte Studien und online Umfragen ausgewertet werden, mit deren Hilfe ähnliche Fragestellungen verfolgt wurden.

Eine solche online Umfrage wird derzeit von DARIAH-DE durchgeführt, wovon erste Resultate in einem Report im Jänner 2015 veröffentlicht wurden (Stiller et al. 2015). Diesem Report ist die

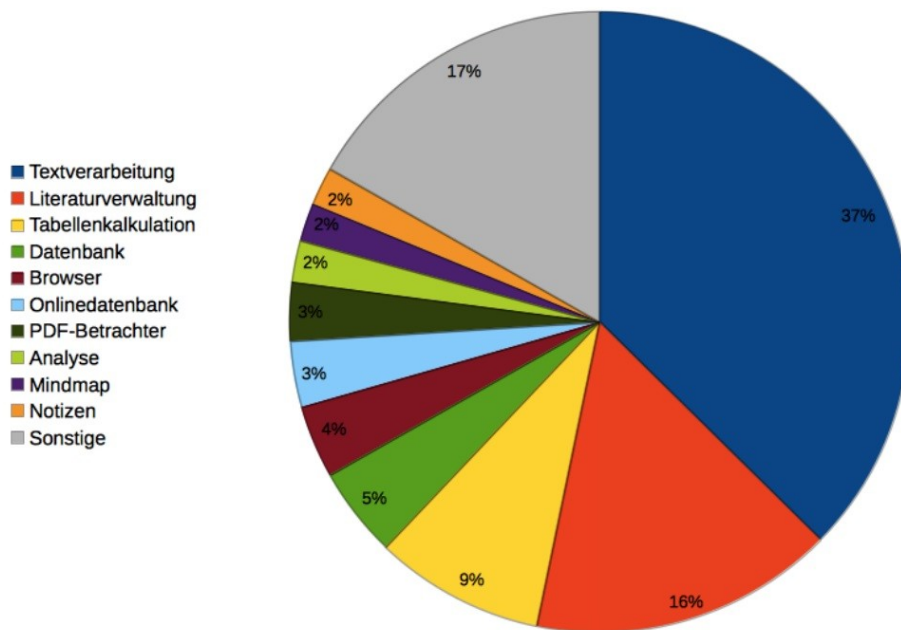
---

4 Der aktuelle Stand der Umfrage kann unter dem folgenden Link abgerufen werden: <http://digital-archiv.at:8081/exist/apps/DARIAH-Collection/pages/InterviewAuswertung.html>.



untenstehende Grafik entnommen, welche ein deutliches Übergewicht in der Verwendung von Textverarbeitungs- und Literaturverwaltungsprogrammen zeigt.

Abbildung 4 Die am häufigsten genannte Software nach Kategorien in allen Phasen



**Abbildung 1: Stiller, Nutzungsverhalten, S. 29; n= 77.**

Eine europaweite und über alle Wissenschaftsbereiche hinweg durchgeführte Umfrage aus dem Jahr 2009 fragte nach den verwendeten „data types“ und förderte das folgende Ergebnis zu Tage (Kuipers and van der Hoeven 2009). Zur Erklärung: mit „Network-based data“ sind Webseiten, Email und „Chat-History“ gemeint (Kuipers and van der Hoeven 2009, S. 30).

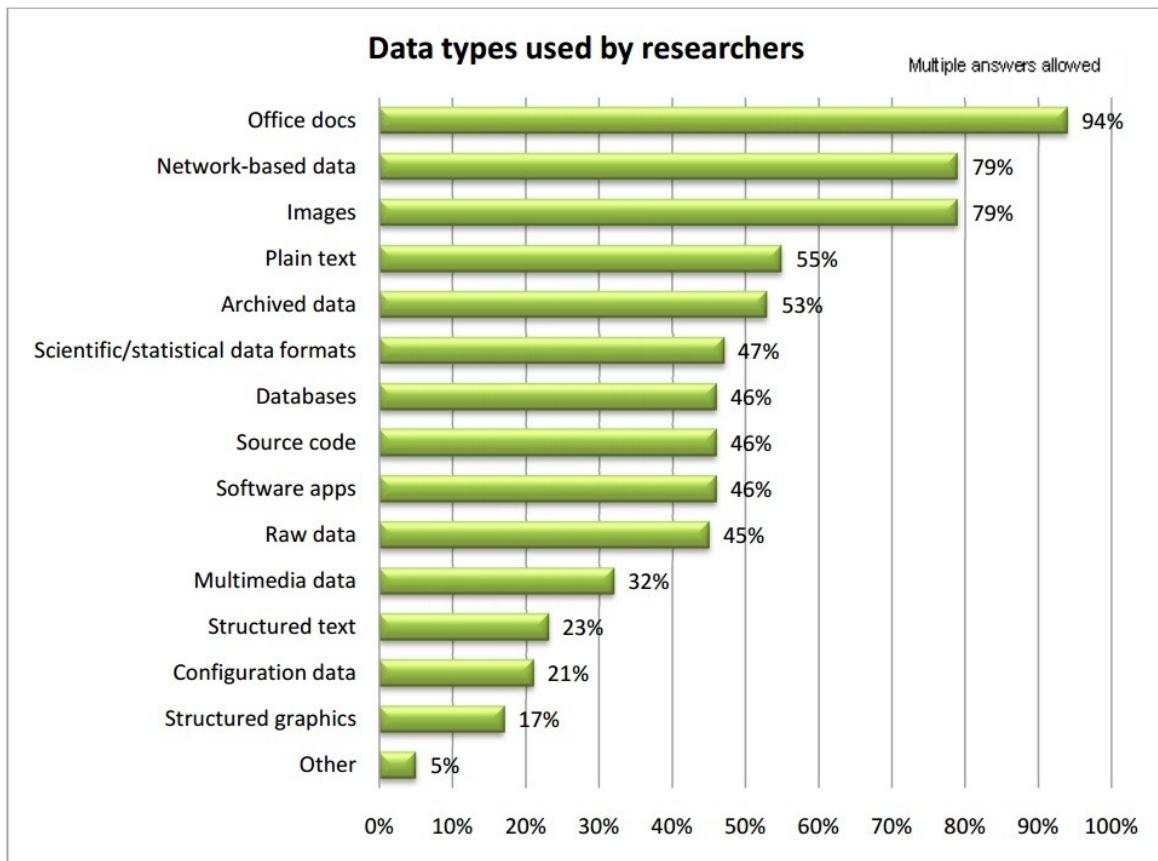


Figure 15: data types used by researchers, n = 1366

**Abbildung 2: Kuipers, PARSE, S. 31; n=1366**

Ein Ergebnis, das von den Autoren der Studie mit folgendem Kommentar versehen wurde: „What is a bit surprising perhaps is that still 6% of the respondents do not use office documents“ (Kuipers and van der Hoeven 2009, S. 31).

Älter noch ist eine von der von den „University of Minnesota Libraries“ im Jahr 2006 durchgeführten Befragung von rund 570 Studenten und Universitätsangehörigen (University of Minnesota Libraries 2006). Auf die Frage „What are your typical research methodologies“ antworteten 84% der „humanists“ mit „textual analysis“. Am seltensten, nämlich gerade mit einmal 3,3 und 5,4 Prozent wurden die Methoden „statistical and quantitative research“ genannt (University of Minnesota Libraries 2006, S. 17 und 18).

Dass die Geisteswissenschaften eine sehr textlastige Wissenschaft sind, geht auch aus einer 2013 an der Berliner Humboldt-Universität durchgeführten Umfrage hervor (Simukovic, Kindler and Schirmbacher 2013). So handelt es sich bei dem klar dominierenden Datentyp um „Texte“.

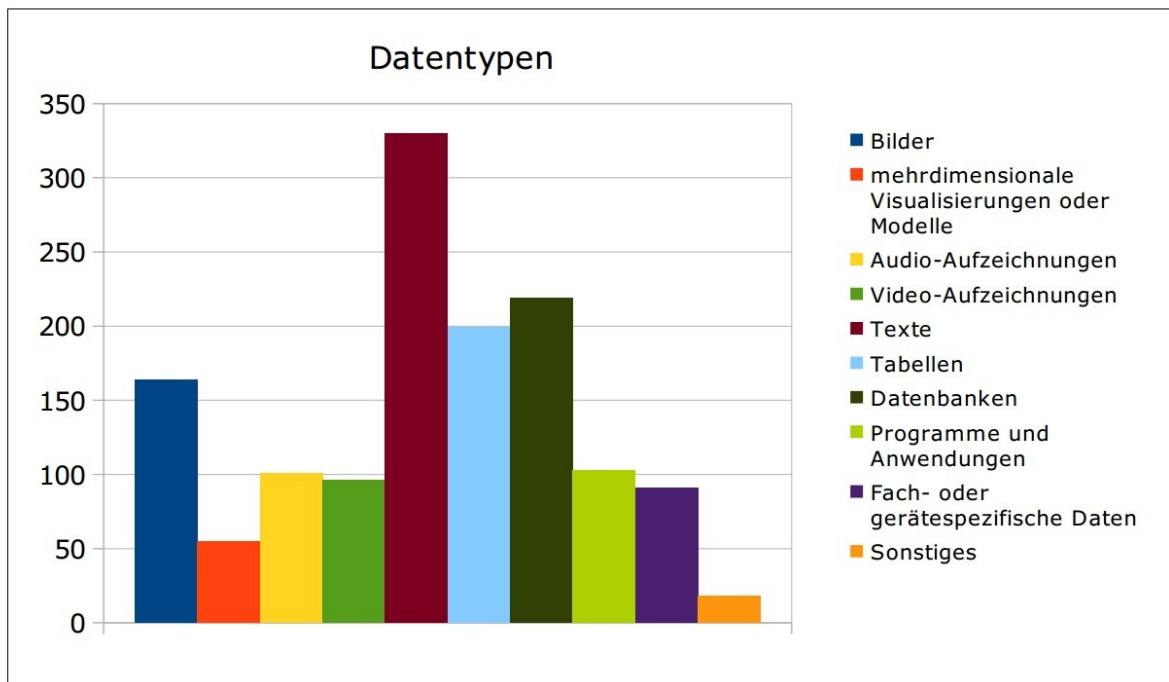


Abbildung 4: Typen von Forschungsdaten

**Abbildung 3: Simukovic, Forschungsdaten, S. 15; n=499.**

Außerdem wurden die Befragten gebeten, ihre „Quellen von Forschungsdaten“ zu identifizieren (Simukovic, Kindler and Schirnbacher 2013, S. 14).

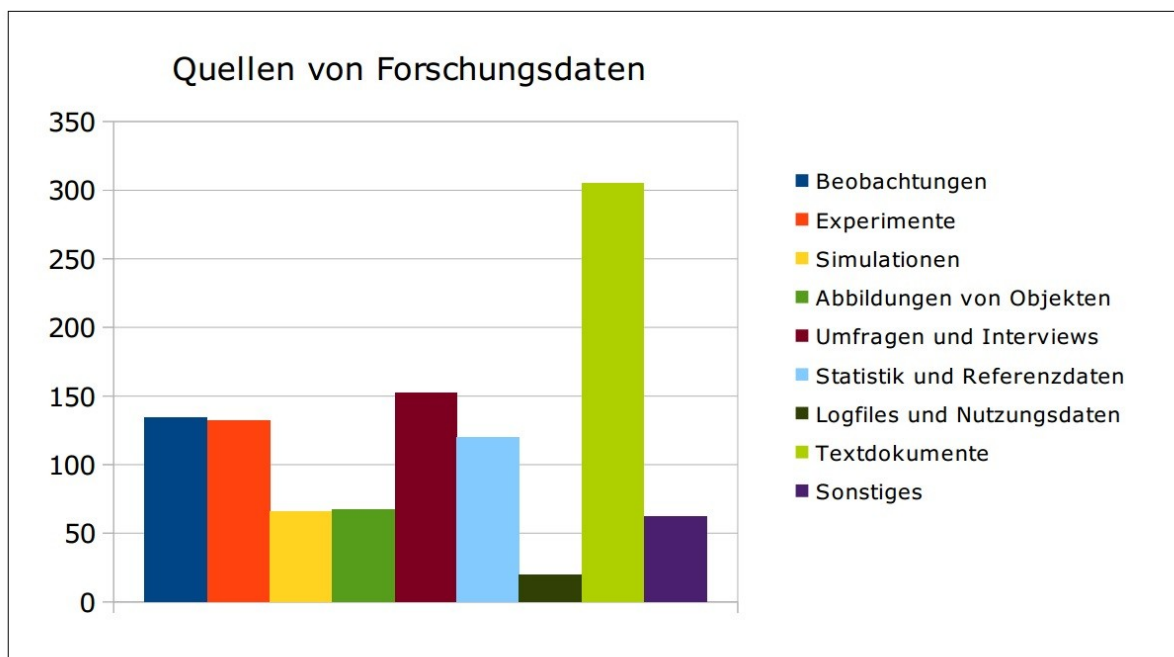


Abbildung 3: Quellen von Forschungsdaten

**Abbildung 4: Simukovic, Forschungsdaten, S. 14; n=499.**

## 3. Forschungsdaten in den Geisteswissenschaften

### 3.1 Spielarten und Definition von Forschungsdaten

Die Gegenüberstellung bzw. die konsequente Abgrenzung der Begriffe „Quellen“ und „Forschungsdaten“ bereitete bei den an der Herzog August Bibliothek durchgeführten Interviews zum Thema Forschen und Forschungsdaten meist große Schwierigkeiten.

Als wenig hilfreich erwies sich dabei die in der Literatur gerne anzutreffende Differenzierung des Forschungsdatenbegriffs in Primär- oder Rohdaten und Sekundärdaten. Bezeichnungen, die auf einen vornehmlich in den Natur- und Sozialwissenschaften gebräuchlichen Forschungsdatenbegriff zurückzuführen sind, also aus jenen Wissenschaftszweigen stammen, die vermehrt mit quantitativen Methoden arbeiten.

Das Verhältnis von Primär- und Sekundärdaten bzw. „Raw“ und „Derived and Recombined Data“, aber auch von Daten und Publikationen, zeigt eine von Jim Gray 2007 vorgestellte „Data Pyramid“ (Hey 2007).

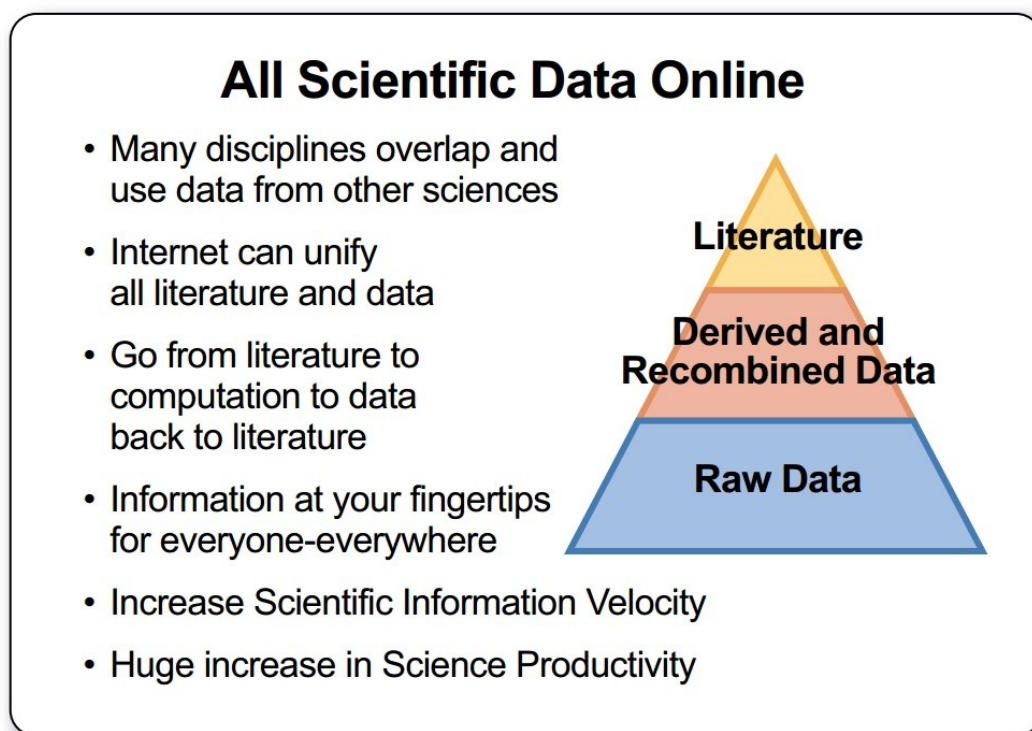
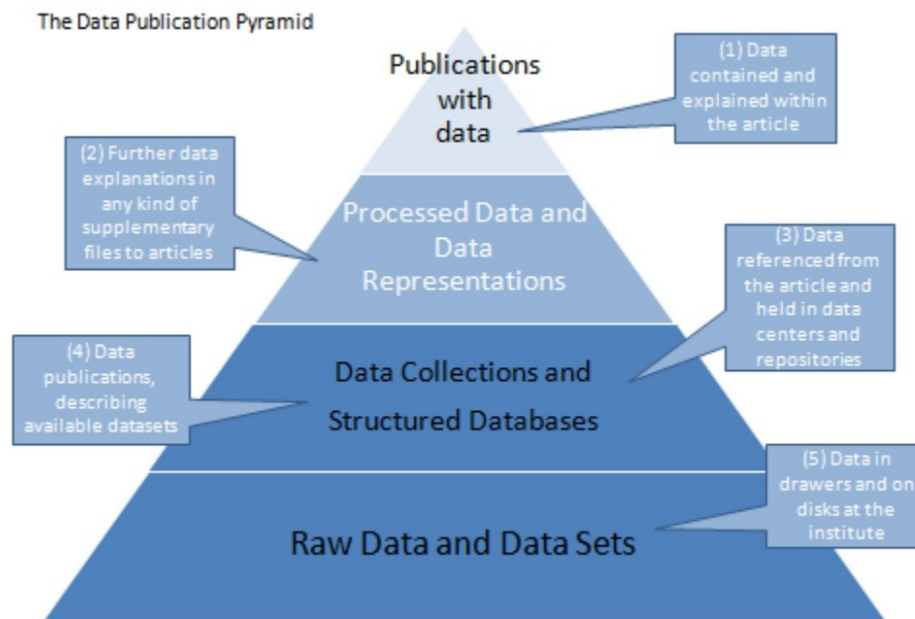


FIGURE 3

Abbildung 5: Hey, Jim Gray on eScience, S. xxvi.

Dieses Modell wurde 2011 von Susan Reilly (u.a.) aufgegriffen und weiterentwickelt (Reilly 2011), wie nachfolgende Grafik zeigt.



Graph 1 repeated, The Data Publication Pyramid, based on the Jim Gray Pyramid.

**Abbildung 6: Reilly, Integration of Data and Publication, S. 19.**

Der Frage, ob man mit diesen vor allem aus den Naturwissenschaften kommenden Termini auch in den Geisteswissenschaften operieren soll oder kann, gingen Hügi und Schneider in ihrer 2013 publizierten Studie zu „Digitale[n] Forschungsinfrastrukturen für die Geistes- und Geschichtswissenschaften“ nach (Hügi and Schneider 2013).<sup>5</sup> Dabei kommen sie zu dem Schluss, dass unter diesem Gesichtspunkt nur noch jene geisteswissenschaftlichen Projekte mit Forschungsdaten operieren würden, „welche eine natur- oder sozialwissenschaftliche Methode benutzt haben, um quantitative Daten zu generieren“ (Hügi and Schneider 2013, S. i.).

Weniger Berührungspunkte, wenigstens was die Terminologie betrifft, zeigen Patrick Sahle und Simone Kronewett. In ihrem ebenfalls 2013 veröffentlichten Aufsatz „Jenseits der Daten. Überlegungen zu Datenzentren für die Geisteswissenschaften am Beispiel des Kölner ‚Data Center for Humanities‘“, unterscheiden die beiden zwischen „Ergebnisdaten auf der einen und Ausgangsdaten oder ‚Primärdaten‘ auf der anderen Seite“ (Sahle and Kronewett 2013, S. 77). In Übereinstimmung mit den zuvor gezeigten Statistiken, entspricht für Sahle und Kronewett nämlich „Text im allerweitesten Sinne“ am ehesten dem Konzept von Primärdaten, mit der Konsequenz, dass „alle unveröffentlichten Quellen, Archivalien, Fundstücke et cetera“, aber explizit auch Sekundärliteratur als Primär- oder auch „Ausgangsdaten“ verstanden werden könnten. Diesen Primär- oder Ausgangsdaten wären dann die in den Geisteswissenschaften traditionellen Formen der Ergebnissicherung und -präsentation gegenüberzustellen. Kurz gesagt die Monographie und der Aufsatz, wofür Sahle und Kronewett die Bezeichnung „Ergebnisdaten“ verwenden (Sahle and Kronewett 2013, S. 77-79).

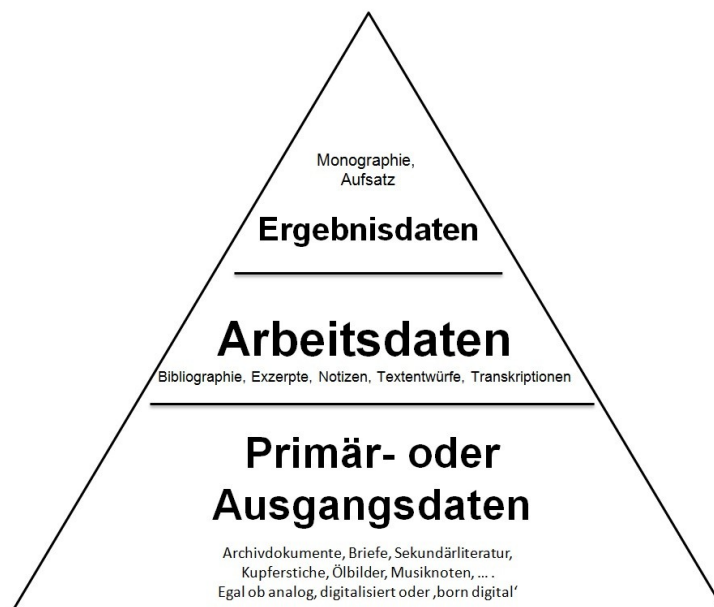
---

<sup>5</sup> Siehe hier v.a. S. 16-20, wo die Autoren u.a. direkt auf Reilly verweisen.

Monographien und Aufsätze schreiben sich jedoch nicht von selbst, sondern sind das Resultat eines teils mehrjährigen Forschungsprozesses bzw. der daraus produzierten Arbeitsdaten. Sahle und Kronewett weiter zitierend:

Tatsächlich entstehen im Forschungsprozess nämlich regelmäßig auch intermediäre oder Arbeitsdaten. In den Projekten werden Materialien gesammelt und erschlossen, Aggregationsstufen erzeugt, Texte bewertet, kommentiert und annotiert, Verlinkungen hergestellt, Aufzeichnungen oder ganze Aktenreihen angelegt, Korrespondenzen geführt, der Arbeitsprozess dokumentiert oder vielfältige analytische oder narrative Zwischenstufen zum endgültigen Ergebnis erarbeitet. In der Literatur ist hier von einer Zwischenschicht zwischen ‚input‘ und ‚output‘ die Rede, die als ‚throughput‘ bezeichnet wird. (Sahle and Kronewett 2013, S. 79)

Verbindet man die von Sahle und Kronewett, zum Teil mit Berufung auf Hügi und Schneider (Hügi and Schneider 2013, S. 79, Fußnote 7) vorgeschlagenen Termini „Primär- oder Ausgangsdaten“, „Intermediär- oder Arbeitsdaten“ und „Ergebnisdaten“ mit der von Gray bzw. Reilly vorgestellten Datenpyramide, so ergibt sich folgende schematische Darstellung:

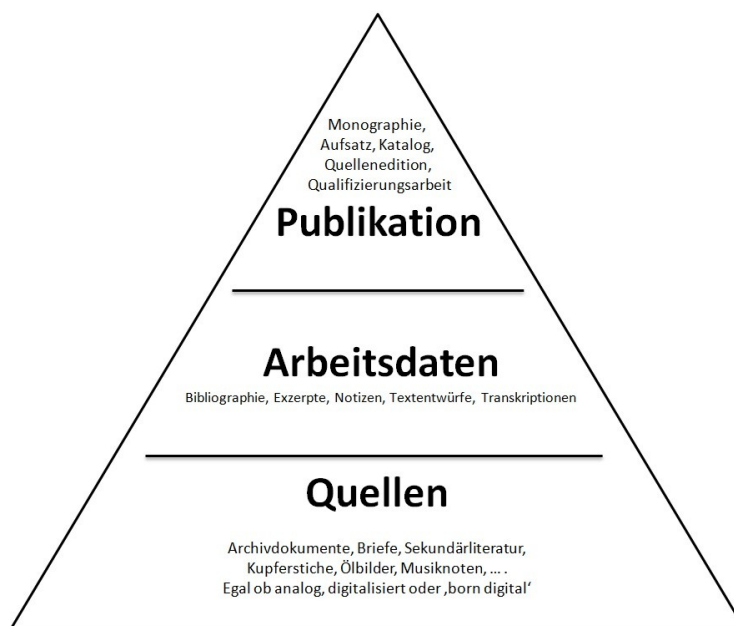


**Abbildung 7: Datenpyramide geisteswissenschaftlicher Forschungsdaten I (P. Andorfer)**

In ihrem Aufsatz fahren Sahle und Kronewett nun fort, diese für die traditionelle geisteswissenschaftliche Forschung „etablierte Struktur“ aus der Perspektive digitaler Geisteswissenschaften kritisch zu hinterfragen. Unter digitalen Geisteswissenschaften wird dabei vor allem die Arbeit mit digitalisierten Quellen, die Arbeit in einer kollaborativen, virtuellen Forschungsumgebung, die Arbeit mit digitalen Werkzeugen und die Publikation von Ergebnissen in „neueren Formen der wissenschaftlichen ‚Mitteilung‘“ wie „Hypertexte[n], Wikis, Blogs und teilweise sehr komplexen Informationsportalen[]“ verstanden (Sahle and Kronewett 2013, S. 80). Vor diesem Hintergrund kommen Sahle und Kronewett zu dem Schluss, dass die „Abtrennbarkeit der Bereiche Primärdaten – Zwischendaten – Ergebnisdaten“ nicht mehr gegeben ist:

Durch die Digitalisierung des Forschungsprozesses verschmelzen die verschiedenen Arten von Forschungsdaten zu einem Kontinuum, das von den Ausgangsdaten bis zu den Narrativen der Ergebnisse der Forschung alle Schritte der Verarbeitung umfasst. (Sahle and Kronenwett 2013, S. 81)

Da es hier in diesem Paper in erster Linie aber um eine Annäherung an den Forschungsdatenbegriff in den traditionellen Geisteswissenschaften geht, kann dieses behauptete „Verschmelzen von Forschungsdaten“ wenigstens vorerst einmal außer Acht gelassen werden. Stattdessen soll hier vielmehr die Frage verfolgt werden, warum bei einem Schema, das eine „etablierte Struktur“ des (geisteswissenschaftlichen) Forschens beschreibt, nicht auch etablierte Begriffe Verwendung finden können. Denn weder Primär-, Ausgangs- oder Rohdaten sind im Begriffsrepertoire traditioneller Geisteswissenschaftlerinnen und Geisteswissenschaftler verankert. Hingegen kann wohl kaum ein Begriff stärker mit geistes- und kulturwissenschaftlichen Disziplinen assoziiert werden, als jener der Quelle. Es empfiehlt sich daher, die in den Geisteswissenschaften eher unüblichen Begriffe Primär-, Ausgangs oder Rohdaten durch den gut eingeführten Begriff „Quelle“ zu ersetzen. Gleiches gilt auch für den Terminus „Ergebnisdaten“, welcher in den Geisteswissenschaften nach wie vor seine Entsprechung in der Monographie, der Qualifizierungsarbeit und im Aufsatz findet. Das Ergebnis einer solchen Begriffs Korrektur zeigt nachfolgende Datenpyramide.



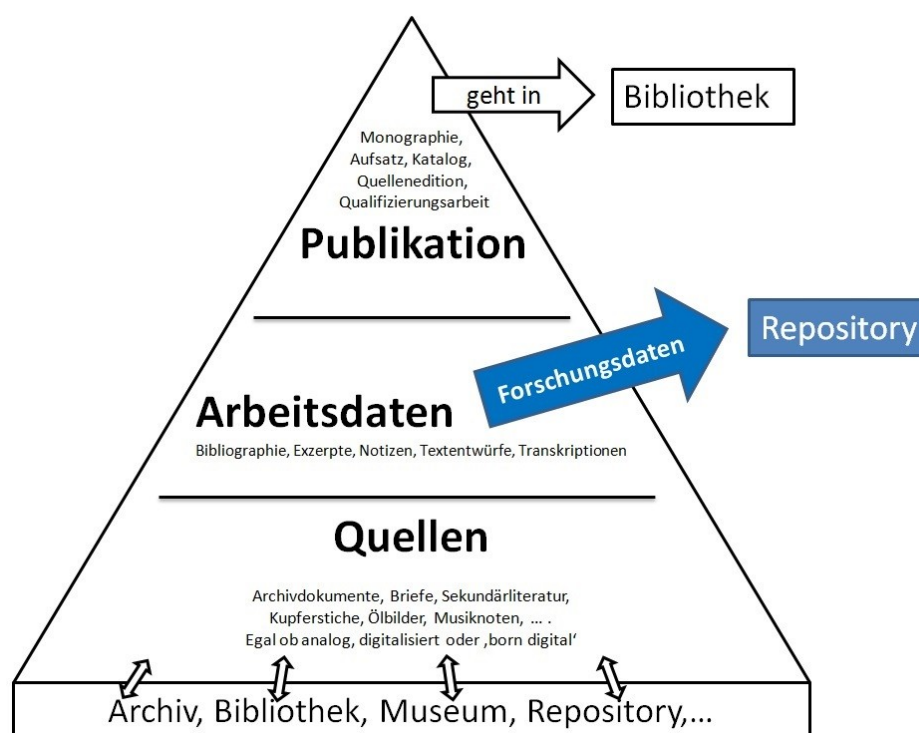
**Abbildung 8: Datenpyramide geisteswissenschaftlicher Forschungsdaten II (P. Andorfer)**

Mit den Begriffen „Quelle“ und „Monographie/Aufsatz“ sind außerdem fest etablierte Institutionen verbunden. Konkret handelt es sich dabei um Archive, Bibliotheken und Museen, welche die Quellen für die Forschung bereitstellen, und zwar gleichgültig ob in analoger oder digitaler Form. Die Veröffentlichung von Ergebnissen geisteswissenschaftlicher Forschungen geschieht wiederum mit Hilfe von Verlagen und Bibliotheken. Diese Institutionen gewährleisten auch die dauerhafte Öffentlichkeit und die Referenzierbarkeit von Quellen und Publikationen. Gerade dieser institutionelle Kontext und die damit einhergehenden Kriterien dauerhafter Öffentlich- und Referenzierbarkeit erweisen sich als ungemein hilfreich in der Annäherung an

eine Definition von geisteswissenschaftlichen Forschungsdaten. Denn im Unterschied zu Quellen oder Publikationen werden die im Laufe eines Forschungsvorhabens erzeugten Arbeitsdaten, wie etwa Transkripte, Exzerpte, Notizen, Vorstufen der Verschriftlichung der Forschungsergebnisse, für den individuellen Gebrauch strukturierte Literatur- und Quellensammlungen, etc. traditionellerweise nicht veröffentlicht. Der damit einhergehende Verlust mag für einen Teil dieser Arbeitsdaten einmal mehr, einmal weniger verschmerzbar sein. Manche Arbeitsdaten hingegen könnten als relevant genug erachtet werden, um als Forschungsdaten in einer eigenen Infrastruktur längerfristig, öffentlich zugänglich und nachnutzbar gespeichert zu werden. Eine Definition von geisteswissenschaftlichen Forschungsdaten könnte daher wie folgt lauten:

*Geisteswissenschaftlich Forschungsdaten sind alle jene Daten,<sup>6</sup> die im Kontext einer geisteswissenschaftlichen Fragestellung und in der Arbeit mit den dabei eingesehenen Quellen, worunter auch Sekundärliteratur verstanden wird, für eine längerfristige und öffentliche Archivierung ausgewählt und aufbereitet werden.*

Folgende Grafik versucht diese Definition eines Forschungsdatenbegriffes in den Geisteswissenschaften schematisch und unter Berücksichtigung des institutionellen Kontextes darzustellen.



**Abbildung 9: Datenpyramide geisteswissenschaftlicher Forschungsdaten III im institutionellen Kontext (P. Andorfer)**

<sup>6</sup> Daten wird hier im Sinne der OAIS-Definition verstanden „Data: A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing“. Vgl. Consultative Committee for Space Data Systems, Reference Model for an Open Archival Information System, Issue 2, Washington 2012, S. 10.



## 3.2 Beispiele geisteswissenschaftlicher Arbeits- und Forschungsdaten

Die Brauchbarkeit der im vorigen Abschnitt entwickelten Definition geisteswissenschaftlicher Forschungsdaten soll im Anschluss nun anhand zweier Beispiele erprobt werden.

### 3.2.1 – Beispiel A

Für das vorliegende Paper wurden knapp 40 Publikationen, vorwiegend Projektberichte und Aufsätze in Zeitschriften und Sammelbänden eingesehen. Sofern die einzelnen Beiträge online zugänglich waren, wurden sie heruntergeladen und in einem eigens für diesen Aufsatz angelegten Dropbox-Ordner gespeichert, wobei versucht wurde, bei der Benennung der einzelnen PDF-Dateien einheitlich vorzugehen. Die Dateien wurden mit dem PDF-Viewer Foxit Reader geöffnet, die Texte gelesen und mit Hilfe von Bookmarks, Hervorhebungen und Kommentaren bearbeitet. All diese individuellen Anmerkungen in den Texten wurden in den PDF-Dateien gespeichert. Aufgrund der genannten Tätigkeiten (systematische Neubenennung, inhaltliche Texterschließung) handelt es sich bei den im Dropbox-Ordner gespeicherten Dateien nicht mehr (nur) um Quellen sondern bereits um Arbeitsdaten. Diese Arbeitsdaten werden noch ergänzt um ein MS-Word Dokument im docx-Format, welches Exzerpte aus besonders relevant erachteten Quellen – den rund 40 Publikationen – versammelt, die ihrerseits wiederum mit eigenen Kommentaren und Notizen versehen sind. Zusätzlich wurden die Titeldaten dieser ausgewählten Publikationen noch in eine eigens für diesen Aufsatz angelegte Zotero-Library eingepflegt. Sowohl die MS-Word Datei, als auch die Zotero-Library stellen ebenfalls Arbeitsdaten dar. Ergänzt wird diese Sammlung von Arbeitsdaten abschließend noch um eine MS-PowerPoint Datei im pptx-Format, da die Abbildungen 7-9 mit MS-PowerPoint erstellt wurden.

Ein erster und technisch, wie vom Arbeitsaufwand einfacher Schritt, um aus den hier beschriebenen Arbeits- auch Forschungsdaten zu machen, wäre die Freigabe und Publikation des Dropbox-Links zu jenem Ordner, in dem die zuvor genannten Dateien gespeichert sind. Damit wäre einmal das Kriterium der Öffentlichkeit der Forschungsdaten erfüllt. Unklar ist jedoch, wie lange diese Daten unter dem veröffentlichten Link wohl erreichbar bleiben werden.<sup>7</sup>

Einer solchen ‚Veröffentlichung‘ müsste, entsprechend der oben vorgestellten Definition von Forschungsdaten aber eine Auswahl der zu publizierenden Arbeitsdaten samt einer nicht näher beschriebenen Aufbereitung vorausgehen. Wie diese Aufbereitung jedoch konkret auszusehen hat bzw. welche Daten als ausreichend relevant zu erachten sind, um überhaupt publiziert zu werden, hängt erstens von der Infrastruktur ab, welche für die Speicherung und Veröffentlichung der Daten zuständig sein soll, zweitens vom jeweiligen Forschungsprojekt und drittens vom intendierten Zweck der Publikation der Forschungsdaten.

Im Falle einer Veröffentlichung der Forschungsdaten via Dropbox hält sich der Arbeitsaufwand für die Datenaufbereitung sehr in Grenzen bzw. fällt zur Gänze weg. Da der Link zu der Forschungsdatensammlung zu diesem Paper nur in diesem Paper selbst publiziert werden würde, würden Forschungsdaten und Publikation zwangsläufig als zusammengehörige Einheit auftreten. Dies hätte einmal zur Konsequenz, dass der Kontext der Forschungsdaten durch die

---

<sup>7</sup> Allfällige rechtliche Aspekte hinsichtlich der Veröffentlichung von Aufsätzen/Artikeln werden hier bewusst ausgeklammert.

Publikation ausreichend tief erschlossen wäre. Dies hätte aber auch zur Konsequenz, dass die Forschungsdaten, losgelöst vom Kontext der Publikation, praktisch ihren gesamten Wert verlieren würden, da aus den Daten alleine weder hervorgeht, wer die einzelnen Dateien gesammelt, bearbeitet und öffentlich zugänglich gespeichert hat, noch ersichtlich ist, zu welchem Zweck und im Rahmen welcher Forschungsfrage diese Daten generiert wurden. Außerdem fehlen jegliche Hinweise bezüglich allfälliger Einschränkungen oder Bedingungen hinsichtlich einer Nachnutzung dieser Daten. Daraus folgt, dass der Arbeitsaufwand für die Datenaufbereitung stark von der Entfernung zum Forschungsprojekt abhängt, in dem die Daten generiert wurden. Dass eine Veröffentlichung von Forschungsdaten in einem kontrollierten Umfeld wie einem Forschungsdatenrepositorium, welches ein Mindestmaß von Metadaten zu einzelnen Forschungsdaten oder auch Forschungsdatensammlungen fordert, helfen kann, die Distanz zwischen Daten und Projekt zu vergrößern, dürfte offensichtlich sein.

Weniger offensichtlich ist jedoch die Beantwortung der Frage, welche Arbeitsdaten überhaupt als Forschungsdaten öffentlich zugänglich und referenzierbar gespeichert werden sollen. Die im Kontext der Diskussion um Forschungsdaten gerne zitierten Empfehlungen des Deutschen Wissenschaftsrates oder der Europäischen Kommission geben als Gründe für die Sicherung und Veröffentlichung von Forschungsdaten die Schlagworte Nachvollziehbarkeit bzw. Überprüfbarkeit sowie Nachnutzbarkeit an (DFG 2013).

Zieht man vorerst einmal nur das Kriterium der Nachvollziehbarkeit von Forschungsergebnissen heran, so ist die Auswahl der zu publizierenden Forschungsdaten rasch getroffen. In Übereinstimmung mit den etablierten Standards für geisteswissenschaftliches Forschen und Schreiben ist die Nachvollziehbarkeit der Ergebnisse durch Fußnotenapparat und Literatur- und ggf. Quellenverzeichnis gewährleistet, wodurch eine Veröffentlichung allfälliger Forschungsdaten überflüssig ist. Zu untersuchen wäre jedoch, inwieweit geisteswissenschaftliche Publikationen bzw. die darin veröffentlichten Ergebnisse allein aufgrund von Quellen- und Literaturangaben tatsächlich nachvollziehbar sind. Zu berücksichtigen wäre hier einerseits der Prozess der Quellenauswahl und -suche und andererseits die Reflexion der mehr oder weniger bewussten Auswahl- und Beurteilungskriterien der Quellen und Literatur. Von den 15 interviewten Geisteswissenschaftlerinnen und Geisteswissenschaftlern an der Herzog August Bibliothek gab nur eine Forscherin an, ihre online-Recherchen (was auch einen Gutteil der Quellen- und Literaturrecherche mit einschließt) systematisch zu dokumentieren und dieses ‚Protokoll‘, also die Suchbegriffe, den Zeitpunkt der Suche und das Suchportal für andere nachvollziehbar in ihrer Publikation zu veröffentlichen. Rund ein Drittel der Befragten gaben an, den Suchprozess in keinerlei Weise zu dokumentieren, während die restlichen zwei Drittel sagten, dies zumindest für sich selbst und meist nur wenig strukturiert zu protokollieren.<sup>8</sup>

Wenn im vorliegenden Beispiel die Nachvollziehbarkeit der Ergebnisse auch ohne Veröffentlichung von Forschungsdaten gegeben ist, würde eine Publikation ausgewählter Arbeitsdaten als Forschungsdaten dann vor dem Hintergrund des Arguments der Nachnutzbarkeit Sinn machen? Diese Frage muss ganz grundsätzlich einmal mit Ja beantwortet werden. So ist es zweifellos bequemer, schneller und praktischer die für diesen Aufsatz gesammelte Literatur zum Thema Forschungsdaten mit einem einzigen Klick aufrufen und herunterladen zu können, als die einzelnen Texte mit Hilfe der jeweiligen in diesem Paper veröffentlichten Hyperlinks Stück für Stück selbst zu öffnen und/oder herunterzuladen. Noch

---

<sup>8</sup> Vgl. dazu: <http://digital-archiv.at:8081/exist/apps/DARIAH-Collection/pages/InterviewAuswertung.html>

besser wäre es freilich, wenn die einzelnen PDF-Dateien nicht nur in Form eines ZIP-Files zum Download zu Verfügung stünden, wie dies etwa bei einer Veröffentlichung via Dropbox der Fall wäre, sondern wenn auf die einzelnen Texte unmittelbar via API zugegriffen werden könnte, wodurch sich diese beispielsweise in einen Topic-Modeling-Workflow integrieren ließen.

Unter dem Gesichtspunkt einer Ökonomisierung des geisteswissenschaftlichen Forschungsprozesses wäre eine Veröffentlichung der hier beschriebenen Arbeitsdaten wohl zu begrüßen. Dies trifft umso mehr auf all jene Literatur und Quellen zu, die noch nicht in digitaler und online zugänglicher Form vorliegen, was wohl die überwiegende Mehrheit der Quellen geisteswissenschaftlichen Forschens umfasst. Dies soll anhand eines zweitens Beispiels demonstriert werden:

### **3.2.1 – Beispiel B**

Ein am Institut für Geschichtswissenschaften und Europäische Ethnologie der Universität Innsbruck angesiedeltes Dissertationsprojekt versucht das Weltbild eines Tiroler Bauern um 1800 zu skizzieren. Grundlage dafür ist eine rund 1000 Seite starke Handschrift, worin ein 1753 in Waidring bei Kitzbühel geborener Bauer die gesamte Welt, deren Regionen, Länder, Geschichte und Bewohner beschreibt (Andorfer 2012). Besagte Handschrift, die „Weltbeschreibung des Leonhard Millinger“ befindet sich im Besitz des Kulturvereins von Kirchdorf, der die Handschrift im eigenen Heimatmuseum ausstellt. Eine Mikrofilmkopie der Handschrift kann außerdem im Tiroler Landesarchiv eingesehen werden (Tiroler Landesarchiv, Mikrofilm Nr. 774, Abschnitt 1). Eine Aufgabe des Dissertationsvorhabens bestand daher darin, diese nur schwer einsehbare Quelle in Form einer digitalen Online-Edition einer breiteren Öffentlichkeit zugänglich zu machen. Außerdem bildet die so aufbereitete Handschrift die Basis für weitere Analysen des Textes wie etwa zum Textaufbau oder den vielfältigen intertextuellen Bezügen zwischen der Handschrift und der vom Verfasser gelesenen Werke. Die digitale Online-Edition wurde als work-in-progress 2013 in der von der Herzog August Bibliothek Wolfenbüttel herausgegebenen Reihe „Editiones Electronicae Guelferbytae“ veröffentlicht und umfasst neben dem in XML-TEI kodierten Text der Weltbeschreibung auch noch eine Einleitung, eine Bibliographie, ein Orts- und Personenregister (allesamt ebenfalls nach XML-TEI kodiert) sowie die Faksimiles der Weltbeschreibung (Andorfer, ed., 2013). Ergänzt wird diese Edition noch um eine eigene Metadatendatei (METS/MODS) und um diverse XSLT-Stylesheets, welche vornehmlich die HTML-Darstellungen realisieren. Die Edition der Weltbeschreibung ist im Katalog der Herzog August Bibliothek verzeichnet und kann somit über eine herkömmliche Literatursuche (z.B. mittels Karlsruher Verbundkatalog) gefunden werden.

Um den Verfasser der Handschrift bestmöglich in seinem sozialen, ökonomischen und kulturellem Umfeld verorten zu können, wurden außerdem umfangreiche Archivrecherchen durchgeführt. Mit Hilfe von Tauf-, Hochzeits- und Sterbematrizen konnte ein mehrere Generationen umfassender Stammbaum erstellt werden, die Durchsicht von Verfachbüchern eröffnete aufgrund der darin überlieferten Hofübergabe- und Erbverträgen Einsichten in das bäuerliche Alltagsleben, eine Auswertung des Grundbuches erlaubte die relative ökonomische Einordnung der Familie Millinger innerhalb der Gemeinde und Visitationsberichte des Vikariats Waidring lieferten Hinweise auf deren soziale Stellung. Die Archivarbeit wurde vornehmlich in Form von Arbeitstranskripten relevant erachteter Quellenfunde dokumentiert, die mit MS-Word angefertigt und als docx-Dokumente gespeichert wurden. Da in den besuchten Archiven weder

Fotografieren noch Scannen gestattet ist und die Erstellung von Kopien oder Scans von Archivalien durch das Archiv sowohl zeit- als auch kostenaufwändig ist, musste in der Regel auf Reproduktionen verzichtet werden. Die Erfassung von klar strukturierten Quellen, wie dem Grundbuch der Gemeinde Waidring, Viehbestandslisten oder Einwohnertabellen, erfolgte mit Hilfe von MS-Excel. Für die Erstellung des Stammbaumes wurde das Programm „Ahnenblatt“<sup>9</sup> verwendet. Die Verwaltung der so erstellten Dateien erfolgt mit Hilfe des Windows-Explorers, wobei für jedes Archiv ein eigener Ordner angelegt wurde, welcher wiederum mehrere, nach Bestandsgruppen geordnete Unterordner enthalten kann.

Was die eingesehene Sekundärliteratur betrifft, so wurde ähnlich wie in Beispiel A vorgegangen, allerdings mit dem Unterschied, dass hier der überwiegende Teil der Literatur nicht frei zugänglich im Internet veröffentlicht ist. Für relevant erachtete Titel wurde daher jeweils ein eigenes MS-Word (docx) Dokument angelegt, in das neben dem Vollzitat auch mehr oder weniger umfangreiche Exzerpte eingetragen wurden. Kleinere Texte wie Aufsätze, Artikel aber auch Abschnitte aus Monographien wurden darüber hinaus teilweise gescannt und als PDF-Dateien gespeichert, vereinzelt erfolgte auch eine Volltexterfassung mittels OCR. Sofern kein Scanner vorhanden war, wurden einige Texte auch abfotografiert und die einzelnen Bilder im jpg-Format in einem jeweils eigenen Ordner gespeichert. Die Anzahl der bisher generierten Dateien liegt bei rund 150 docx-Dateien (ca. 10 MB) mit Exzerpten und knapp 100 Digitalisaten (ca. 3 GB).

Um allfällige sprachliche Besonderheiten in Leonhard Millingers Schriften entdecken und analysieren zu können, wurde ein Korpus vergleichbarer schriftlicher Zeugnisse angelegt. Dieses setzt sich abgesehen von den Texten Millingers aus Transkriptionen von Quellen zusammen, die von Kollegen für deren eigene Projekte angefertigt und von diesen dankenswerterweise in digitaler Form (docx) zur Verfügung gestellt wurden (Span 2014). Für die Untersuchung wurden die einzelnen Texte als txt-Dateien gespeichert und mit Hilfe von AntConc sowie mit eigenen Python- und R-Skripten untersucht. Zu weiteren Analysen und Visualisierungen der Weltbeschreibung des Leonhard Millinger wurden die XML-TEI Dateien der Edition in eine eXist-db Datenbank geladen.

Die Synthese der Ergebnisse bzw. deren Verschriftlichung erfolgte in MS-Word. Die Erstellung des Literaturverzeichnisses soll durch Zotero unterstützt werden, wofür die bibliographischen Angaben der verwendeten Sekundärliteratur in eine eigene Zotero-Bibliographie eingepflegt wurden.

Sämtliche der hier genannten Dateien sind in einem eigenen Verzeichnis auf der lokalen Festplatte bzw. aufgrund der Verwendung des Microsoft Cloud-Dienstes OneDrive auch in der Wolke gespeichert, abgesehen von der digitalen Edition der Weltbeschreibung, deren Dateien auf einem Server der Herzog August Bibliothek liegen. Doch bei welchen dieser Dateien handelt es sich nun um Quellen, um Arbeitsdaten, um Forschungsdaten oder um Publikationen?

Wie schon in Beispiel A handelt es sich bei den im Laufe des Forschungsprozesses angefertigten Transkripten, Exzerpten, Bibliographien und Digitalisaten von Quellen und Literatur allesamt um Arbeitsdaten. Arbeitsdaten, die beinahe per Knopfdruck – im konkreten Fall ist damit die Freigabe des entsprechenden OneDrive-Ordners samt Veröffentlichung des dazugehörigen Links gemeint – zu Forschungsdaten werden könnten. Ebenfalls in Analogie zu Beispiel A stellt

---

<sup>9</sup> <http://www.ahnenblatt.de/>.

sich aber auch hier die Frage nach der Notwendigkeit der Veröffentlichung dieser Daten. Aus Gründen der Nachvollziehbarkeit der in der Dissertation zu publizierenden Ergebnisse wäre – siehe Beispiel A – eine Veröffentlichung nicht notwendig, da die Quellen und die Literatur, aus denen diese Arbeitsdaten generiert wurden, in der Publikation angeführt werden und zumindest in der Theorie von jedem selbst eingesehen werden können.

Im Unterschied zu Beispiel A ist die Einsichtnahme in die verwendeten Quellen und Literatur aber ungleich aufwendiger, einerseits aufgrund der häufigen Standortgebundenheit der Quellen und andererseits aufgrund methodischer Hindernisse wie etwa die benötigten paläographischen Kenntnisse. Dass aufgrund dieser Hürden eine Veröffentlichung von Forschungsdaten zu besonders schwer zugänglichen Quellen die Nachnutzbarkeit derselben drastisch erhöht, sollte aus diesem Beispiel deutlich geworden sein.

Weniger deutlich dürften jedoch die im Falle einer Veröffentlichung derartiger Daten zu berücksichtigenden rechtlichen Implikationen sein. Da dieses Thema aber den Rahmen vorliegender Arbeit sprengen würde, muss hier auf eine weitere Verfolgung dieser komplexen Materie verzichtet werden. Nicht verzichtet werden soll jedoch auf die Verortung der erwähnten digitalen Edition der Weltbeschreibung und der für Analysezwecke eingerichteten eXist-db Datenbank innerhalb der oben skizzierten Datenpyramide. Da es sich dabei aber weniger um Produkte traditioneller geisteswissenschaftlicher Forschung handelt, wird darauf erst im nächsten Abschnitt, der sich dezidiert mit Forschungsdaten in den Digital Humanities auseinandersetzt, eingegangen werden.

Zuvor gilt es noch ein knappes Fazit hinsichtlich des Begriffes Forschungsdaten in den (traditionellen) Geisteswissenschaften zu ziehen.

### **3.3 Ein Fazit**

Sofern kein massiver Umbruch des traditionellen geisteswissenschaftlichen Verständnisses von wissenschaftlicher Nachvollziehbarkeit der eigenen Forschungsergebnisse angestrebt werden soll, bedarf es in den allermeisten Fällen weder des Begriffs der geisteswissenschaftlichen Forschungsdaten, noch Mittel und Wege zur Publikation selbiger. Denn die Nachvollziehbarkeit der Ergebnisse sollte durch die (formalisierte) Offenlegung der verwendeten Quellen und der konsultierten Literatur (Fußnotenapparat, Quellen- und Literaturverzeichnis) samt einer verschriftlichten Begründung der Auswahl der Quellen und Literatur gegeben sein.

Wenn es aber darum gehen soll, Ergebnisse besser, sprich im Sinne von Nachnutzbarkeit zu publizieren, so ist dies wohl erst mit Hilfe des Begriffes geisteswissenschaftlicher Forschungsdaten möglich.

## 4. Forschungsdaten in den digitalen Geisteswissenschaften

### 4.1 Digitale Geisteswissenschaften

Die anhaltenden Debatten um die Frage nach dem Wesen, den Eigenschaften und Spezifika der digitalen Geisteswissenschaften bzw. der Digital Humanities ist bekannt, sollen hier aber nicht weiter referiert werden (Huggett 2012).<sup>10</sup> Im Kontext dieser Arbeit wird unter digitalen Geisteswissenschaften einerseits die Generierung und Veröffentlichung maschinenlesbarer Derivate geisteswissenschaftlicher Quellen sowie andererseits die Auseinandersetzung mit geisteswissenschaftlichen Fragestellungen mit vornehmlich computergestützten, quantitativen Methoden verstanden, wobei ersteres als Material für zweiteres dient – sofern nach der Phase der Datenaufbereitung noch Ressourcen für eine weitere Datenverarbeitung vorhanden sind (Thaller 2012, S. 10). Dabei soll aber nicht verschwiegen werden, dass diese Perspektive besonders den Tätigkeitsbereich der digitalen Editorik forciert.<sup>11</sup>

Diese Schwerpunktsetzung lässt sich in zweierlei Hinsicht rechtfertigen. Ohne repräsentative Zahlen nennen zu können (Grupe and Kerzel 2015, S. 277, Abbildung 6), scheint derzeit gerade das Erstellen von digitalen Editionen das Brot- und-Butter-Geschäft der digitalen Geisteswissenschaften zu sein, wobei hier die Bezeichnung digitale Edition auch die vielfältigen und meist in Form von Datenbanken erschlossenen Bild-, Objekt- und Datensammlungen (z.B. zu Personen) miteinbezieht. Diese (behauptete) quantitative Dominanz digitaler Editorik scheint auch insofern plausibel zu sein, da regelbasierte und maschinenlesbare Quellen – die „Ergebnisse“ digitaler Editorik – die Voraussetzung für eine weitere computergestützte Analyse darstellen.

Diese Schwerpunktsetzung erweist sich vor allem aber auch im Zusammenhang mit dem Forschungsdatenbegriff als sinnvoll, gilt es doch die Frage zu stellen, wo bzw. auf welcher Ebene in der im vorigen Abschnitt vorgestellten Datenpyramide (vgl. Abbildung 9) eine digitale Edition zu verorten ist bzw. ob eine solche klare Zuordnung überhaupt sinnvoll und hilfreich ist.

Die Frage nach der Verortung von digitalen Editionen und Datenbanken in einer Datenpyramide lässt sich dann eindeutig beantworten, wenn davon ausgegangen wird, dass Sinn und Zweck einer digitalen Edition oder Datenbank in der Erschließung, Aufbereitung, vor allem aber in der Veröffentlichung verschiedenster ausgewählter Materialien liegen. Eine digitale Edition ist somit in erster Linie als eine abgeschlossene und öffentlich zugängliche Einheit zu betrachten, die wie eine Publikation Endpunkt eines Forschungsprojektes ist und wie eine Quelle Ausgangspunkt eines weiteren Forschungsprojektes sein kann. Von vorneherein als zu publizierendes Ergebnis angelegt, unterscheidet sich die digitale Edition somit von jenen Arbeits- bzw. Forschungsdaten, die im Laufe der Auseinandersetzung mit einer konkreten geisteswissenschaftlichen Fragestellung und auf dem Weg zu einer Publikation (Qualifizierungsarbeit, Aufsatz, Monographie) generiert werden und deren Publikation nicht das primäre Ziel des Forschungsvorhabens war.

---

<sup>10</sup> Für einen ausführlichen Literaturüberblick zur Frage nach dem Wesen der Digital Humanities siehe außerdem Terras, Nyhan, and Vanhoutte, *Defining Digital Humanities: A Reader*, Ashgate 2013.

<sup>11</sup> Für eine ausgefeiltere Definition siehe Thaller, „Controversies around the Digital Humanities: An Agenda“, S. 11f.

Diese Einschätzung entspricht auch jener Christine Borgmans, welche hinsichtlich „simulations, data repositories, and other complex content with interactive links“ konstatiert, dass diese am ehesten als Publikationen zu betrachten wären („may [be] considered publications“), ohne diese Aussage aber näher zu begründen (Borgman 2010, S. 99).

Die Analogien zwischen digitalen und analogen Editionen verschiedenster Quellen und Publikationen können außerdem anhand zweier konkreter Beispiele festgemacht werden. Dafür sei einmal auf die in Beispiel B erwähnte digitale online Edition der Weltbeschreibung des Leonhard Millinger verwiesen. Diese erschien in einer von der Herzog August Bibliothek herausgegebenen Reihe, wobei die Bibliothek auch die permanente Zugänglichkeit und Referenzierbarkeit der Edition garantiert. Außerdem erhöht die Aufnahme der Edition im Katalog der Bibliothek deren Sichtbarkeit. Auf formaler bibliographischer Ebene können in diesem Falle somit keine nennenswerten Unterschiede zwischen einer gedruckten und einer digitalen Quellenedition beobachtet werden.<sup>12</sup>

Parallelen zwischen analogen und digitalen Edition eröffnen sich auch hinsichtlich des höchst traditionellen Prozesses der Qualitätssicherung geisteswissenschaftlicher Publikationen – dem Verfassen und Veröffentlichen von Rezensionen. Zumindest seit Juni 2014, dem Erscheinungsmonat der ersten Ausgabe von „ride“, einem vom Institut für Dokumentologie und Editorik herausgegebenen *review journal for digital editions and resources*, verfügen auch „digital editions and resources“ über ein eigenes Medium für Rezensionen.<sup>13</sup>

Als weniger eindeutig erweist sich die Frage nach der Verortung von digitalen Editionen und Datenbanken jedoch aus einer institutionellen Perspektive. Vereinfacht formuliert: Quellen findet man im Archiv, in der Bibliothek und im Museum. Publikationen oder Literatur findet man in Bibliotheken. Um ihre Ressourcen für die interessierte Öffentlichkeit auffindbar und dauerhaft zugänglich zu machen, haben die genannten Einrichtungen gut etablierte Systeme (Findbuch, Katalog) ausgebildet. Digitale Editionen und Datenbanken können nun Teil dieser Systeme sein, wie das Beispiel der Edition der Millinger'schen Weltbeschreibung zeigt, oder eben auch nicht. Bei einer Suche im Karlsruher Verbundkatalog (KVK) nach den im *review journal for digital editions and resources* besprochenen digitalen Editionen konnten gerade einmal drei von insgesamt zehn Editionen gefunden werden.<sup>14</sup> Damit soll aber nicht behauptet werden, dass digitale Ressourcen wie Editionen nicht oder nur schwer gefunden werden können, sofern sie nicht in die traditionellen Nachweissysteme eingetragen werden. Eine Google-Suche mit den selben Suchstrings, wie sie bei der Recherche im KVK verwendet wurden, listete jede der rezensierten Editionen unter die Top-Drei Ergebnissen. Diese Beobachtung soll vielmehr als Argument dafür dienen, dass digitale Editionen, wenigstens aus Perspektive der Bibliotheken nicht als sammlungs- und katalogisierungswürdige Objekte bzw. Publikationen angesehen werden.

---

12 Vgl. dazu etwa die OPAC-Einträge zur Edition der Millinger'schen Weltbeschreibung (<http://opac.lbs-braunschweig.gbv.de/DB=2/XMLPRS=N/PPN?PPN=77097337X>) und zum 10. Band der Edition der Nuntiaturreportagen Deutschlands (<http://opac.lbs-braunschweig.gbv.de/DB=2/XMLPRS=N/PPN?PPN=71666755X>).

13 Siehe <http://www.i-d-e.de/> und <http://ride.i-d-e.de/>.

14 Gesucht wurde nach dem Titel bzw. einem Teil des Titels (im Titel-Feld) sowie dem Erscheinungsjahr. Gefunden wurden der „Codex Sinaiticus“ (Nachweis in der „Israel Union List“), „Sandrat.net“ (Nachweis im Union Catalog Hesse) sowie das „Shelly Goldwin Archive“ (Nachweis in WoldCat).

Die Konsequenz daraus ist klar. Ohne einer eindeutigen und dauerhaften institutionellen Verankerung von digitalen Editionen und Datenbanken ist eine über die Projektlaufzeit hinausgehende Veröffentlichung von digitalen Editionen und Datenbanken nicht gesichert, ein Unterschied zu den auf der eigenen Festplatte liegenden Arbeitsdaten nicht gegeben, da eine persistente Referenzierung nicht möglich ist.

Eine Veröffentlichung digitaler Editionen und Datenbanken analog zur Publikation geisteswissenschaftlicher Forschungsdaten wie im vorigen Kapitel definiert, kann dabei aber nur bedingt als Lösung betrachtet werden. Denn während solche geisteswissenschaftlichen Forschungsdaten tatsächlich mehrheitlich voneinander unabhängige Einzeldateien sind, verfügen digitale Editionen und Datenbank meist über einen die Summe der Einzeldateien überschreitenden und häufig auch funktionellen Mehrwert. Ein Mehrwert, der beispielsweise durch (verschiedene) Darstellungs-, Analyse- und Abfragemöglichkeiten generiert wird. Gerade dieser Mehrwert dürfte jedoch dafür verantwortlich sein, dass sich traditionelle Institutionen wie eben Bibliotheken und Archive gegenwärtig (noch) kaum für digitale Editionen und Datenbanken verantwortlich fühlen, weil es für eine längerfristige Verfügbarhaltung komplexerer digitaler Ressourcen auch technischer und personeller Ressourcen bedarf.

Aus zukünftiger infrastruktureller und institutioneller Perspektive kann eine Unterscheidung zwischen komplexeren digitalen Ressourcen, also digitalen Editionen oder Datenbanken einerseits, und geisteswissenschaftlichen Forschungsdaten andererseits, somit relativ einfach getroffen werden. Da bei zweiteren kein nennenswerter, sprich über die in den Forschungsdaten gespeicherte Information hinausgehender Mehrwert vorliegt, ist die dauerhafte Speicherung mit relativ geringem Aufwand möglich. Erstere hingegen, die sich oftmals erst über Funktionalität und Zusammenwirken vieler Einzelteile definieren, stellen eine ungleich größere Herausforderung dar, will man sie längerfristig und in möglichst unveränderter Form und Funktionalität zugänglich halten.

## 4.2 Repository und Datenzentrum

Die im vorigen Abschnitt angestellten Überlegungen zur institutionellen Verortung von digitalen Editionen und Datenbanken und den damit einhergehenden Schwierigkeiten sind weder neu noch originell, sondern schon länger bekannt und häufig diskutiert, etwa auch auf der Dhd-Tagung in Graz 2015. Auf dieser präsentierte sich die „AG Datenzentren im Verband DHd“<sup>15</sup> mit einem Panel zum Thema „Was sind und was sollen Datenzentren in den Geisteswissenschaften“ (Centre for Information-Modelling 2015). Die Agenda dieser Datenzentren kann dem Book of Abstracts zur Tagung entnommen werden:

Datenzentren gewährleisten die Zugänglichkeit und die Nutzbarkeit geisteswissenschaftlicher Forschungsdaten jenseits der aktiven Projektlaufzeit. Aufgaben der Datenkuration (z.B. Dokumentation des Projektkontextes, Migration von Datenformaten) werden in enger Zusammenarbeit mit Wissenschaftlern durchgeführt. Existierende Werkzeuge der Langzeitarchivierung decken dabei nicht immer die spezifischen Anforderungen *geisteswissenschaftlicher Forschungsdaten* ab; so entstehen z.B. erst langsam Ansätze für die Langzeitverfügbarkeit von *Datenbanken, Präsentationssystemen und interaktiven Visualisierungen*. (Centre for Information-Modelling 2015, S. 188)<sup>16</sup>

---

15 <http://www.dig-hum.de/arbeitsgruppe-datenzentren>

16 Eigene Hervorhebung durch Kursivsetzung



Obwohl eine elaborierte Definition von „geisteswissenschaftlichen Forschungsdaten“ hier fehlt – die Diskussion des Forschungsdatenbegriffes war ein Ziel des Panels – wird in der eben zitierten Passage ein sehr deutlicher Forschungsdatenbegriff transportiert. Demnach liegt der Fokus dieser Datenzentren nicht auf jener Art von Forschungsdaten, wie sie in diesem Paper zu definieren versucht werden, sondern auf komplexeren Ressourcen wie „Datenbanken, Präsentationssystemen und interaktiven Visualisierungen“, wobei diese Liste getrost noch um digitale Editionen erweitert werden darf.<sup>17</sup> Dabei steht außer Streit, dass es diesen Ressourcen derzeit an institutioneller Zugehörigkeit mangelt. Außer Streit müsste aber auch stehen, dass sich Datenbanken, digitale Editionen aber auch „Präsentationssysteme und interaktive Visualisierungen“ von den in Kapitel 3 beschriebenen Arbeits- und Forschungsdaten unterscheiden, vor allem was den Grad der intendierten Veröffentlichung und die für eine solche Veröffentlichung benötigten technischen Rahmenbedingungen betrifft.

Nicht umsonst wurde in den beiden Beispielfällen die Publikation ausgewählter Arbeitsdaten via Dropbox oder OneDrive angeregt. Denn einmal unabhängig von rechtlichen Bedenken, entspricht ein solch niederschwelliges und für die einzelne Geisteswissenschaftlerin, für den einzelnen Geisteswissenschaftler, problemlos zu handhabendes System dem Idealtyp eines Forschungsdatenrepositoriums. Ein Repositorium, das eine „fire and forget“ Lösung anbietet, sprich ausgewählte Arbeitsdaten werden mit einem minimalen Arbeitsaufwand einmal ins System eingespeist, wo sie dann für einen mittelfristigen Zeitraum und je nach Belieben und Bedarf von anderen heruntergeladen und weiterverwendet werden dürfen.

Von einem solchen Forschungsdatenrepositorium (wie etwa dem DARIAH-DE Repository) wäre ein Datenzentrum, wie von der AG Datenzentren konzipiert, deutlich zu unterscheiden und zwar deshalb, weil die Ressourcen in einem Datenzentrum – wie eine Publikation – primär dafür generiert wurden, dass andere Personen in Zukunft damit arbeiten können, weshalb auch für eine längerfristige Verfügbarkeit gesorgt werden muss. Anders als bei Forschungsdaten in einem Repository bedarf es bei einem Datenzentrum aber einer ständigen Kuration der Ressourcen um deren Funktionalität auch unter veränderten technischen Voraussetzungen gewährleisten zu können.

---

17 Vgl. dazu die vom auf der Homepage des Digital Humanities Centre, einem Mitglieder der AG Datenzentren im Verband Dhd, angeführten Referenzprojekte [http://humanities-data-centre.org/?page\\_id=220](http://humanities-data-centre.org/?page_id=220). Drei der insgesamt fünf genannten Projekte führen die Bezeichnung „Edition“ direkt im Projekttitel.

## 5. Fazit

Für die weitere Diskussion lässt sich aus eben Gesagtem die klare Empfehlung ableiten, eine präzisere Terminologie zu verwenden. Die bisher verbreitete Praxis, bei der der Begriff geisteswissenschaftliche Forschungsdaten so gut wie alles bedeuten konnte, egal ob es sich um ein Arbeitstranskript im doc-Format, ein von der Bayerischen Staatsbibliothek veröffentlichtes Digitalisat eines alten Druckes, oder eine im Rahmen eines befristeten Projektes erstellte Online-Edition eines Briefwechsels handelt, führt nämlich nicht nur zu unnötiger Verwirrung, sondern birgt auch die Gefahr eines inflationären Gebrauchs dieses Begriffes und in letzter Konsequenz zu einer unnötigen Abwertung des dahinterstehenden Konzeptes.

Dies betrifft weniger jene komplexeren Ressourcen, die zukünftig hoffentlich in Datenzentren ihre institutionelle Heimat finden werden. Zu befürchten ist vielmehr, dass ein Überstrapazieren des Begriffes Forschungsdaten, noch dazu bei einer unpräzisen Verwendung, gerade in der Kommunikation zwischen den digitalen und den traditionellen Geisteswissenschaften eine Kluft erzeugt bzw. bereits (oder noch) bestehende Gräben vergrößert. Vor allem dann, wenn mangels fehlender definitorischer Schärfe der Eindruck entsteht, dass der Begriff Forschungsdaten in erster Linie die in den Geisteswissenschaften fest etablierten Konzepte von Quelle und Literatur/Publikation ersetzen soll. Ein Eindruck, der insbesondere dann entsteht, wenn versucht wird, diese Konzepte als Primär-, Sekundär- oder gar Tertiärdaten zu beschreiben. So liegt dann der Verdacht nahe, dass die Einführung des Begriffes der Forschungsdaten in den Geisteswissenschaften vornehmlich dazu dient, sich zumindest terminologisch den Naturwissenschaften anzunähern, oder – um eine pointierte Formulierung zu verwenden – anzubiedern.

Der wissenschaftliche Mehrwert, der sich daraus ergibt, dass man Quellen künftig etwa Primärdaten und Publikationen von Forschungsergebnissen vielleicht Tertiär- oder Forschungsdaten nennt, dürfte mehr als überschaubar ausfallen. Der wissenschaftliche Mehrwert jedoch, der sich daraus ergibt, dass man all jene Materialien, die während eines Forschungsprozesses generiert werden, nun Arbeitsdaten nennt und einige dieser Arbeitsdaten in Form von Forschungsdaten öffentlich zugänglich und nachnutzbar macht, birgt hingegen das Potential, die über Jahrhunderte kaum veränderte Tradition geisteswissenschaftlichen Forschens nachhaltig zu ändern, und zwar sowohl in quantitativer als auch in qualitativer Hinsicht. Durch die Veröffentlichung geisteswissenschaftlicher Forschungsdaten – zusätzlich zu den traditionellen Publikationen (Monographie, Aufsatz, Qualifizierungsarbeit, Quelleneditionen, Ausstellungskatalogen, ...) – steigert sich die Menge geisteswissenschaftlichen Outputs, wobei dieser Output von anderen Forscherinnen und Forschern, für andere Projekte und für andere Fragestellungen nachgenutzt werden kann. Der Gang ins Archiv, ins Museum oder in die Bibliothek wird dadurch in vielen Fällen zwar nicht ersetzt (zum Glück), kann aber sicherlich besser vorbereitet werden, mit der Konsequenz, dass geisteswissenschaftliches Forschen effektiver wird.

Abgesehen von diesen naheliegenden Effekten, die vornehmlich Aspekte der Zusammenarbeit und Nachnutzbarkeit in den Vordergrund rücken, eröffnet das Konzept geisteswissenschaftlicher Forschungsdaten aber auch die Chance, die häufig nur schwer zu dokumentierenden und damit auch nicht immer zur Gänze nachvollziehbaren Ergebnisse

hermeneutisch-interpretativer Forschungsergebnisse transparenter zu gestalten. Dies kann exemplarisch anhand des vorliegenden Papers demonstriert werden. Die darin zitierten und paraphrasierten Berichte und Aufsätze werden im anschließenden Literaturverzeichnis aufgelistet. Nicht aufgelistet werden darin jedoch jene Texte, welche in Vorbereitung dieses Papers in Augenschein genommen, aber nicht direkt für diese Arbeit verwendet wurden. Eine Veröffentlichung einer solchen erweiterten Bibliographie würde das vorliegende Paper allerdings stärker im individuellen Wissenshorizont des Verfassers kontextualisieren. Würde eine Leserin, ein Leser dieses Papers dann einen bestimmten Aufsatz, ein spezielles Buch in dieser Literaturliste vermissen, so wäre klar, dass das besagte Werk dem Verfasser vermutlich nicht bekannt war. Befände sich ein solcher Aufsatz jedoch nur auf der erweiterten Bibliographie, so kann daraus geschlussfolgert werden, dass dieser Text vom Verfasser als weniger essentiell eingestuft wurde, als die im Verzeichnis der zitierten und paraphrasierten Titel genannten Titel.

Zu bedenken ist außerdem, dass die Ergebnisse der Literaturrecherche für diesen Aufsatz vornehmlich durch den Download der einzelnen Paper bzw. durch das Anlegen von MS-Word Dokumenten mit Exzerpten dokumentiert wurden. Da alle diese Dateien in einem Dropbox-Ordner auf der lokalen Festplatte gespeichert wurden, kann anhand des dabei jeweils automatisch angelegten Datensatzes „Erstellungsdatum“ sehr präzise der zeitliche Ablauf dieser Literaturrecherche rekonstruiert werden, wobei sich aus diesem Wissen wiederum Hinweise auf Zitationsketten ergeben, oder sich der Stellenwert einzelner Titel für diese Arbeit erklären und relativieren lässt.<sup>18</sup>

Obwohl für dieses Paper nicht zutreffend, sei außerdem noch auf Versionierungs- oder Verlaufsprotokolle hingewiesen. Wäre dieser Text etwa mit Google Docs geschrieben worden, so könnte anhand der Verlaufsgeschichte minutiös rekonstruiert werden, wann was geschrieben oder verändert wurde. Da anstelle von Google Docs aber mit LibreOffice gearbeitet wurde, kann anhand der „Dokumentstatistik“ nur eine kumulative Bearbeitungszeit bzw. aufgrund der dokumentierten „Entstehungszeit“ und „Änderungsdatum“ der Bearbeitungszeitraum bestimmt werden.

Durch die Veröffentlichung von Forschungsdaten können solche also nicht nur von anderen nachgenutzt werden, sie erleichtern auch die Nachvollziehbarkeit von Forschungsergebnissen für andere, da sie etwa Informationen zum zeitlichen Ablauf des Literatur- und Quellenstudiums sowie zum größeren Wissenshorizont der Verfasser vermitteln können. Gleichzeitig bleibt zu hoffen, dass die konkrete Beschäftigung mit Arbeits- und Forschungsdaten zu einer vermehrten Reflexion der eigenen Forschungspraxis führt, was sich ebenfalls positiv auf die Qualität der Forschungsergebnisse auswirken dürfte.

---

<sup>18</sup> Selbiges trifft natürlich auch bei einer Verwendung von Literaturverwaltungsprogrammen wie Zotero oder Citavi zu.

## 6. Literaturverzeichnis

- Andorfer, Peter, ed. *Die Weltbeschreibung des Leonhard Millinger*. Editiones Electronicae Guelferbytanæ 12. Wolfenbüttel: Herzog August Bibliothek, 2013.
- . "Die „Weltbeschreibung des Leonhard Millinger‘ als Schlüssel zum „Weltbild‘ des mitteleuropäischen Bauern? Viele Fragen und einige Antworten." In *Forschungswerkstatt: Die Habsburgermonarchie Im 18. Jahrhundert*, edited by Gunda Barth-Scalmani, 225–40. Jahrbuch der Österreichischen Gesellschaft zur Erforschung des achtzehnten Jahrhunderts, 2012.
- . *Forschen und Forschungsdaten in den Geisteswissenschaften: Zwischenbericht einer Interviewreihe*. DARIAH-DE Working Papers 10. Göttingen: Niedersächsische Staats- und Universitätsbibliothek Göttingen, 2015. <http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2015-3-8>.
- Behrens, Julia, Lars Fischer, Karl-Heinz Minks, and Lena Rösler. "Die internationale Positionierung der Geisteswissenschaften in Deutschland. Eine empirische Untersuchung," 2010.
- Borgman, Christine L. *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge, Mass.: MIT Press, 2010.
- Burdick, Anne. *Digital Humanities*. Cambridge, Mass. [u.a.]: MIT Press, 2012.
- Centre for Information-Modelling, and Austrain Centre for Digital Humanities. *Book of Abstracts. DHd2015. Von Daten zu Erkenntnissen. 23. bis 27. Februar 2015, Graz*. Graz, 2015.
- DFG. *Vorschläge zur Sicherung guter Wissenschaftlicher Praxis*. Ergänzte Auflage. Bonn, 2013. [http://www.dfg.de/download/pdf/dfg\\_im\\_profil/reden\\_stellungnahmen/download/empfehlung\\_wiss\\_praxis\\_1310.pdf](http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf).
- Europäische Kommission. "Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. Version 16 December 2013," 2013. [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf).
- Grupe, Nadja, and Martina Kerzel. "TextGrid in Zahlen. NutzerInnen, MitarbeiterInnen, Finanzen." In *TextGrid: Von Der Community - Für Die Community. Eine Virtuelle Forschungsumgebung Für Die Geisteswissenschaften*, edited by Heike Neuroth, Andrea Rapp, and Sybille Söring, 271–85. Göttingen, 2015. [http://webdoc.sub.gwdg.de/univerlag/2015/TextGrid\\_book.pdf](http://webdoc.sub.gwdg.de/univerlag/2015/TextGrid_book.pdf).
- Hey, Tony. "Jim Gray on eScience: A Transformed Scientific Method. Based on the Transcript of a Talk given by Jim Gray to the NRC-CSTB in Mountain View, CA, on January 11," 2007.
- Huggett, Jeremy. "Core or Periphery? Digital Humanities from an Archaeological Perspective." *Historical Social Research* 141, no. 37/3 (2012): 86–105.
- Hügi, Jasmin, and René Schneider. "Digitale Forschungsinfrastrukturen für die Geistes- und Geschichtswissenschaften." Genf: Haute École de Gestion de Genève, 2013. [http://www.infoclio.ch/sites/default/files/standard\\_page/studie\\_forschungsinfrastrukturen\\_small.pdf](http://www.infoclio.ch/sites/default/files/standard_page/studie_forschungsinfrastrukturen_small.pdf).
- Kuipers, Tom, and Jeffrey van der Hoeven. "Insight into Digital Preservation of Research Output in Europe (PARSE). Survey Report," 2009. [http://www.parse-insight.eu/downloads/PARSE-Insight\\_D3-4\\_SurveyReport\\_final\\_hq.pdf](http://www.parse-insight.eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq.pdf).
- Kurz, Susanne. *Digital Humanities: Grundlagen und Technologien für die Praxis*. Wiesbaden: Springer Vieweg, 2015.

- Lauer, Gerhard. "Die digitale Vermessung der Kultur. Geisteswissenschaften als Digital Humanities." In *Big Data*, by Heinrich Geiselberger and Tobias Moorstedt, 99–116. Berlin, 2013.
- Oberhofer, Andreas, and Andreas Hofer. *Weltbild eines "Helden": Andreas Hofers schriftliche Hinterlassenschaft*. Innsbruck: Wagner, 2008.
- Pempe, Wolfgang. "Geisteswissenschaften." In *Langzeitarchivierung von Forschungsdaten. Eine Bestandsaufnahme.*, edited by Heike Neuroth, 137–59. Göttingen, 2012.  
[http://nestor.sub.uni-goettingen.de/bestandsaufnahme/nestor\\_lza\\_forschungsdaten\\_bestandsaufnahme.pdf](http://nestor.sub.uni-goettingen.de/bestandsaufnahme/nestor_lza_forschungsdaten_bestandsaufnahme.pdf).
- Reilly, Susan. "Report on Integration of Data and Publications," 2011.  
<http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/10/ODE-ReportOnIntegrationOfDataAndPublications.pdf>.
- Sahle, Patrick, and Simone Kronenwett. "Jenseits der Daten. Überlegungen zu Datenzentren für die Geisteswissenschaften am Beispiel des Kölner ‚Data Center for the Humanities‘." *Libreas* 23 (2013): 76–96.
- Simukovic, Elena, Maxi Kindling, and Peter Schirmbacher. "Forschungsdaten an der Humboldt Universität zu Berlin. Bericht über die Ergebnisse der Umfrage zum Umgang mit digitalen Forschungsdaten an der Humboldt-Universität zu Berlin," 2013.
- Span, Michael. "Ein Bürger unter Bauern? Michael Pfurtscheller und das Stubaital 1750 bis 1850." Innsbruck, 2014.
- Stiller, Juliane, Klaus Thoden, Oona Leganovic, Christian Heise, Mareike Höckendorff, and Timo Gnad. "Nutzungsverhalten in den Digital Humanities," 2015.  
<https://dev2.dariah.eu/wiki/download/attachments/26150061/Report1.2.1-final.pdf?version=6&modificationDate=1430303390972&api=v2>.
- Terras, Melissa, Julianne Nyhan, and Edward Vanhoutte, eds. *Defining Digital Humanities: A Reader*. Farnham [u.a.]: Ashgate, 2013.
- Thaller, Manfred. "Controversies around the Digital Humanities: An Agenda." *Historical Social Research*. Cologne : GESIS, Leibniz Inst. for the Social Sciences, 1979-, 141, no. 37/2 (2012): 7–23.
- University of Minnesota Libraries. "A Multi-Dimensional Framework for Academic Support. A Final Report," 2006. [http://www2.lib.umn.edu/about/mellon/UMN\\_Multi-dimensional\\_Framework\\_Final\\_Report.pdf](http://www2.lib.umn.edu/about/mellon/UMN_Multi-dimensional_Framework_Final_Report.pdf).