



# Forschungsdaten in den Geisteswissenschaften

Versuch einer Konkretisierung

Peter Andorfer

DHd-Tagung Graz, 25. Februar 2015

# Agenda

1. Einleitung
2. Geisteswissenschaften und deren Methoden
  - 2.1 Geisteswissenschaften der Wissenschaftspolitik
  - 2.2 Qualitative und quantitative Erhebungen zum Status-Quo der Geisteswissenschaften
3. Spielarten und Definition von Forschungsdaten
4. Ein Beispiel und eine Vision (wenn die Zeit reicht)

# 2.1 Geisteswissenschaften der Wissenschaftspolitik

The screenshot shows a Foxit Reader window displaying a PDF document. The left sidebar contains a table of contents with the following items:

- Langzeitarchivierung von Forschungsdaten
- Inhaltsverzeichnis
- Vorwort
- 1 Langzeitarchivierung von Forschungsdaten
- 2 Stand der Diskussion und Aktueller Stand
- 3 Vorgehensweise
- 4 Sozialwissenschaften
- 5 Psycholinguistik
- 6 Pädagogik und Erziehungswissenschaften
- 7 Geisteswissenschaften
  - 7.1 Einführung in den Forschungsbereich
  - 7.2 Kooperative Strukturen
  - 7.3 Daten und Metadaten
  - 7.4 Interne Organisation
  - 7.5 Perspektiven und Visionen
- 8 Altertumswissenschaften
- 9 Geowissenschaften
- 10 Klimaforschung
- 11 Funktionelle Biodiversität
- 12 Medizin
- 13 Teilchenphysik
- 14 Astronomie und Astrophysik
- 15 Zusammenfassung und Internationale Zusammenhänge
- 16 Erkenntnisse und Thesen zur Langzeitarchivierung
- Literaturverzeichnis
- Verzeichnis von Abkürzungen und Symbole
- Autorenverzeichnis

The main content area shows the following text:

## 7 Geisteswissenschaften

*Wolfgang Pempe*

### 7.1 Einführung in den Forschungsbereich

Das räumliche, zeitliche, thematische und methodologische Spektrum der Disziplinen, die unter dem Sammelbegriff der Geisteswissenschaften zusammengefasst werden, ist so vielfältig<sup>1</sup>, dass eine klare Eingrenzung des Forschungsbereichs schwerfällt, zumal die These der „zwei Kulturen“<sup>2</sup> und damit eine deutliche Trennung zwischen Geistes- und Naturwissenschaften längst nicht mehr haltbar ist.<sup>3</sup> In einigen Teilbereichen verschwimmen die Grenzen – und immer häufiger lässt sich eine, u.a. durch „digitale“ Methoden beförderte, interdisziplinäre Zusammenarbeit beobachten, z.B. wenn es um die Altersbestimmung oder Analyse von Objekten oder die Rekonstruktion mit dem menschlichen Auge nicht (mehr) wahrnehmbarer Schriftzeugnisse geht. Aber auch naturwissenschaftliche Disziplinen bedienen sich geisteswissenschaftlicher Methoden, z.B. im Zusammenspiel von Linguistik und Bioinformatik<sup>4</sup> oder in den Neurowissenschaften<sup>5</sup>. **Daneben ergeben sich durch die Anwendung aus Sicht mancher traditioneller Disziplinen neuartiger Methoden, wie z.B. quantitativer Analysen, neue Forschungsfelder.** Bezogen auf den Bereich der Forschungsdaten bedeutet

1 Als ein Beispiel für viele sei hier die Definition des Wissenschaftsrats angeführt: „Philosophie, die Sprach- und Literaturwissenschaften, die Geschichtswissenschaften, die Regionalstudien, die Religionswissenschaften, die Ethnologie sowie die Medien-, Kunst-, Theater- und Musikwissenschaften“ (Wissenschaftsrat (2006), S. 17).

2 Snow (1987).

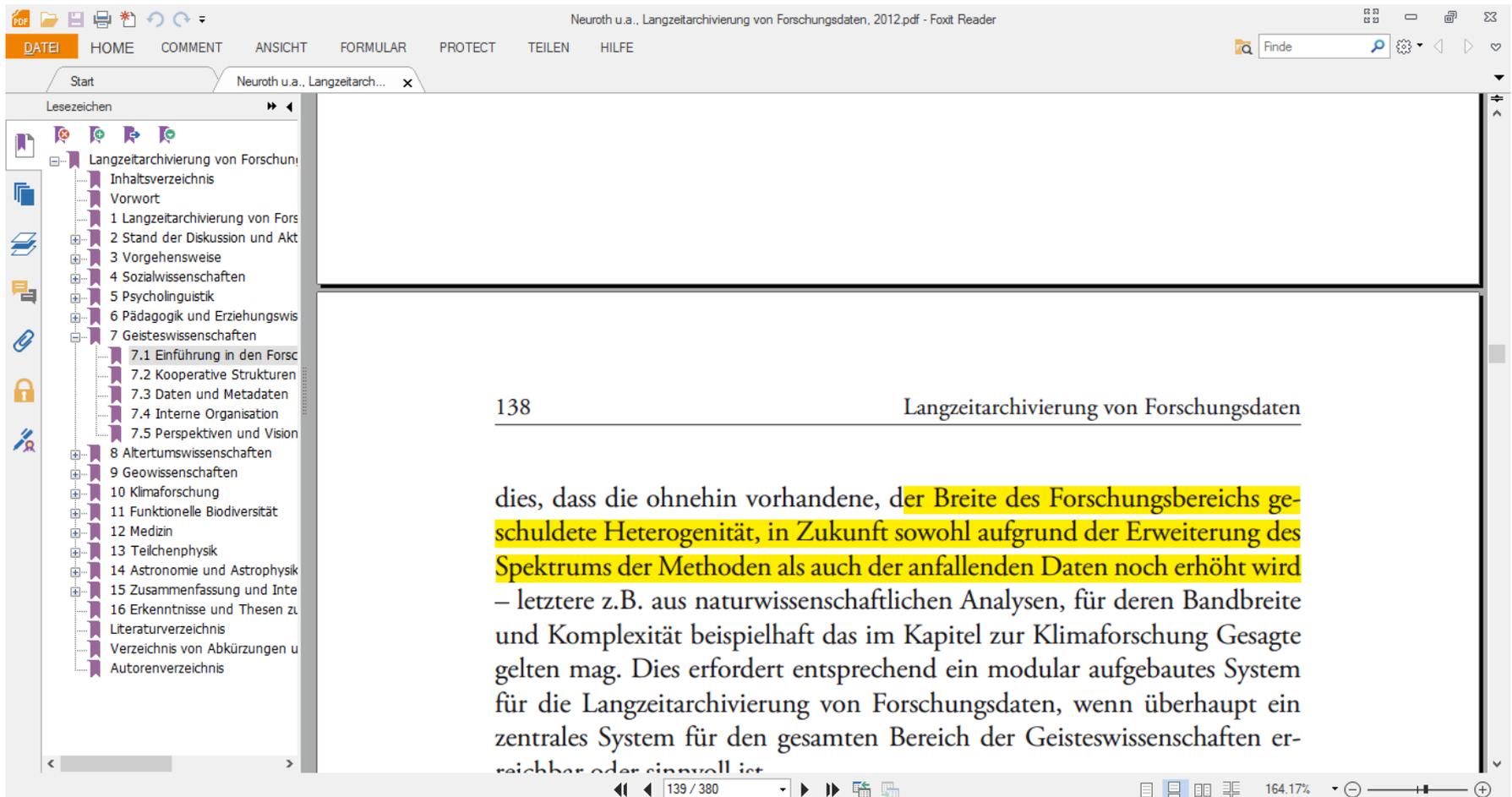
3 Vgl. hierzu Behrens et al. (2010), Kap. 1.

4 „Wechselwirkungen zwischen linguistischen und bioinformatischen Verfahren, Methoden und Algorithmen: Modellierung und Abbildung von Varianz in Sprache und Genomen“ (BMBWF (2011g)).

5 Vgl. z.B. das Projekt Brain/Concept/Writing (humtec (2011)) oder etwas allgemeiner WissensLogs (2011).

The bottom of the window shows a navigation bar with page number 138 / 380, a search bar, and a zoom level of 75.00%.

# 2.1 Geisteswissenschaften der Wissenschaftspolitik



The screenshot shows a Foxit Reader window displaying a PDF document titled "Neuroth u.a., Langzeitarchivierung von Forschungsdaten, 2012.pdf". The left sidebar shows a table of contents with the following items:

- Langzeitarchivierung von Forschungsdaten
  - Inhaltsverzeichnis
  - Vorwort
  - 1 Langzeitarchivierung von Forschungsdaten
  - 2 Stand der Diskussion und Aktueller Stand
  - 3 Vorgehensweise
  - 4 Sozialwissenschaften
  - 5 Psycholinguistik
  - 6 Pädagogik und Erziehungswissenschaften
  - 7 Geisteswissenschaften
    - 7.1 Einführung in den Forschungsdaten
    - 7.2 Kooperative Strukturen
    - 7.3 Daten und Metadaten
    - 7.4 Interne Organisation
    - 7.5 Perspektiven und Visionen
  - 8 Altertumswissenschaften
  - 9 Geowissenschaften
  - 10 Klimaforschung
  - 11 Funktionelle Biodiversität
  - 12 Medizin
  - 13 Teilchenphysik
  - 14 Astronomie und Astrophysik
  - 15 Zusammenfassung und Internationale Zusammenarbeit
  - 16 Erkenntnisse und Thesen zur Langzeitarchivierung
  - Literaturverzeichnis
  - Verzeichnis von Abkürzungen und Zeichen
  - Autorenverzeichnis

The main content area shows page 138, titled "Langzeitarchivierung von Forschungsdaten". The text on the page reads:

dies, dass die ohnehin vorhandene, der Breite des Forschungsbereichs geschuldete Heterogenität, in Zukunft sowohl aufgrund der Erweiterung des Spektrums der Methoden als auch der anfallenden Daten noch erhöht wird – letztere z.B. aus naturwissenschaftlichen Analysen, für deren Bandbreite und Komplexität beispielhaft das im Kapitel zur Klimaforschung Gesagte gelten mag. Dies erfordert entsprechend ein modular aufgebautes System für die Langzeitarchivierung von Forschungsdaten, wenn überhaupt ein zentrales System für den gesamten Bereich der Geisteswissenschaften erreichbar oder sinnvoll ist.

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

The screenshot shows a PDF viewer interface. At the top, there is a menu bar with options: DATEI, HOME, COMMENT, ANSICHT, FORMULAR, PROTECT, TEILEN, HILFE. A search bar on the right contains the text 'Finde'. Below the menu, the document title 'Lauer, Geisteswissenschaf...' is visible. On the left, a sidebar titled 'Seiten' displays four thumbnails of document pages, numbered 1, 2, and 3. The main viewing area shows a text document with several paragraphs. The first paragraph is highlighted in yellow and reads: 'Der Namenswechsel markiert nicht einfach den nächsten »turn«, man reagiert damit vielmehr auf ein Problem, das die Wertehierarchie der kulturwissenschaftlichen Disziplinen betrifft. Das Rechnen gehörte lange Zeit so gut wie nicht zum Methodeninventar der Geisteswissenschaften (vgl. dazu auch den Beitrag von danah boyd und Kate Crawford in diesem Band), wobei methodische Innovationen in diesem Bereich ohnehin eine eher untergeordnete Rolle spielen. In den Natur- und Lebenswissenschaften ist der Einsatz von Computern längst selbstverständlich; an spezialisierte Disziplinen wie Computational Physics, Bio-, Geo- und andere Bindestrich-Informatiken hat man sich gewöhnt; die Vergabe des Physiknobelpreises an Gerd Binnig und Heinrich Rohrer, die 1986 für die Entwicklung des Rastertunnelmikroskops ausgezeichnet wurden, belegt den Stellenwert methodischer Innovationen in diesen Fächern.' The second paragraph is also highlighted in yellow and reads: 'In den Geisteswissenschaften ist das anders. Ihre historisch-hermeneutischen Verfahren werden auf die hellenistische Textkritik oder den Humanismus, auf die Altphilologie eines Christian Gottlob Heyne oder Friedrich August Wolf sowie Philosophen des 19. Jahrhunderts zurückgeführt, man denke etwa an Wilhelm Dilthey. Die Tradition adelt die humanistischen Disziplinen, und wenn sich doch einmal der Wunsch nach Modernität und Innovation regt, wird dieser durch die theoretischen »turns« hinreichend bedient. So weiß man sich stets à jour, und kulturkritische Urteilsroutinen über den Gegensatz von kalter, technischer Zivilisation und warmer, ver-'. The bottom of the viewer shows navigation controls, a page indicator '2 / 18', and a zoom level of '125.00%'.

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

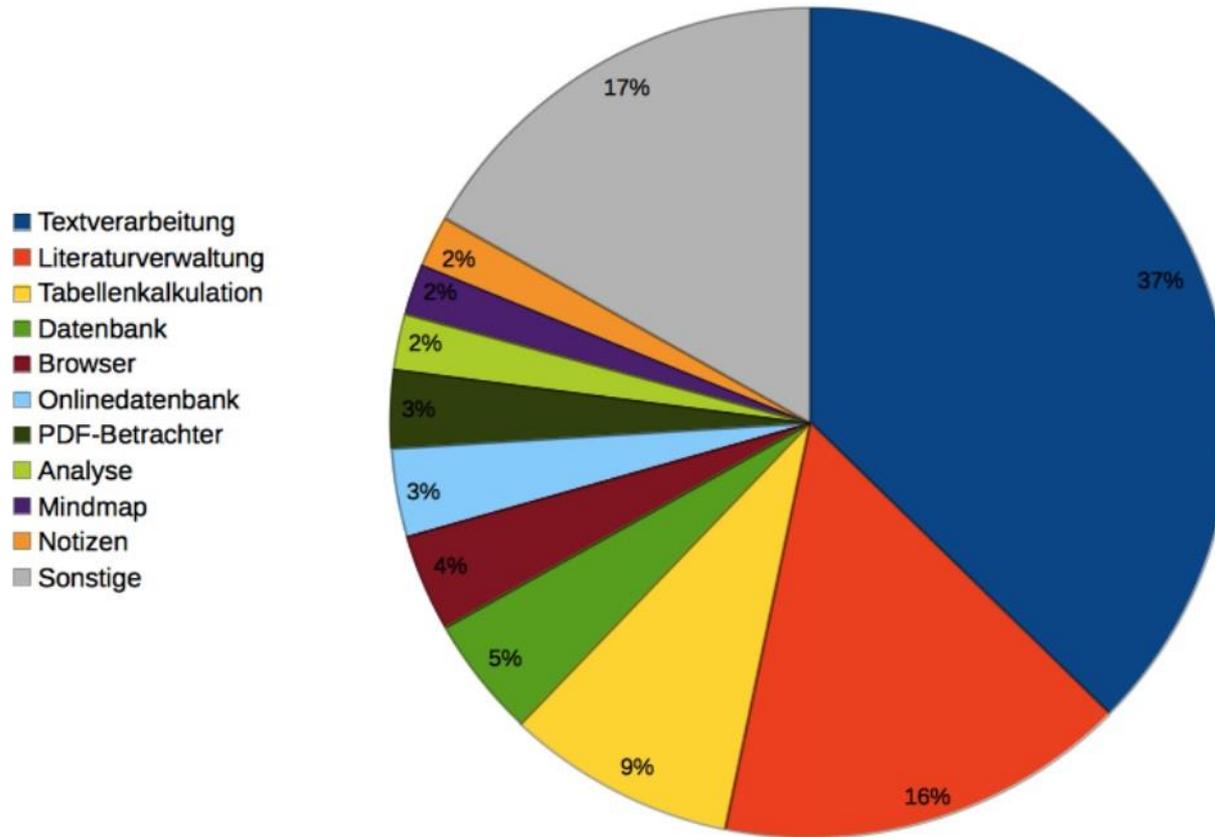
The screenshot shows a Google search for "gerhard lauer". The search results include a snippet for "gerhard lauer :: Home" with the URL "gerhardlauer.de/". Below this, there are image thumbnails for "Bilder zu gerhard lauer". To the right, a knowledge panel for "Gerhard Lauer" is displayed, featuring a portrait and a description: "Gerhard Lauer ist ein deutscher Literaturwissenschaftler und Professor für Deutsche Philologie an der Universität Göttingen. Wikipedia". Below the knowledge panel, there is a section "Wird auch oft gesucht" with five suggested names: Matias Martínez, Wolfgang Frühwald, Peter Strohsch..., Klaus Weimar, and Eric Donald Hirsch. At the bottom of the search results, there is a link to "Georg-August-Universität Göttingen - Lauer, Gerhard, Prof. Dr." with the URL "www.uni-goettingen.de/de/11812.html".

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

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

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

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



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

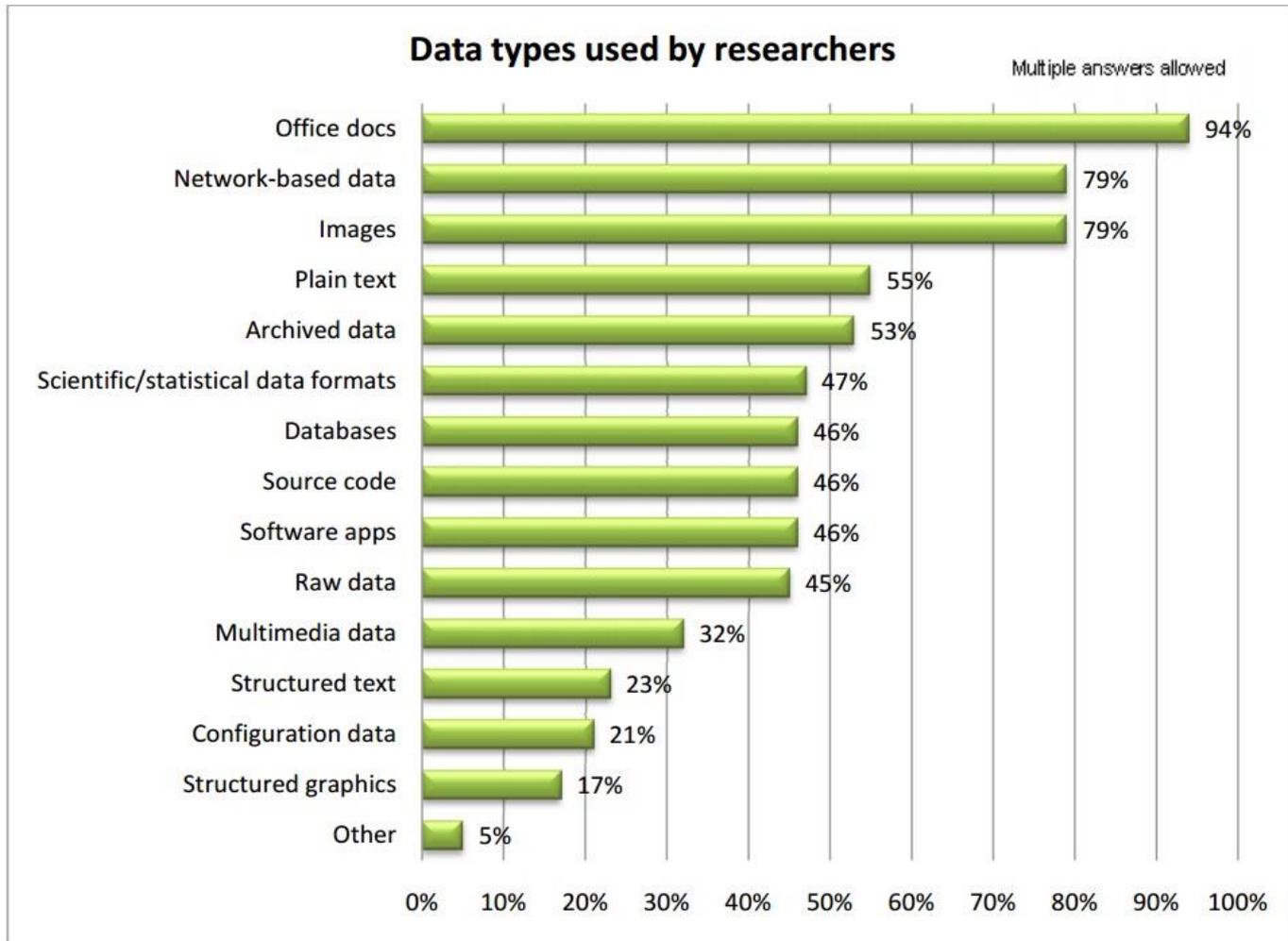
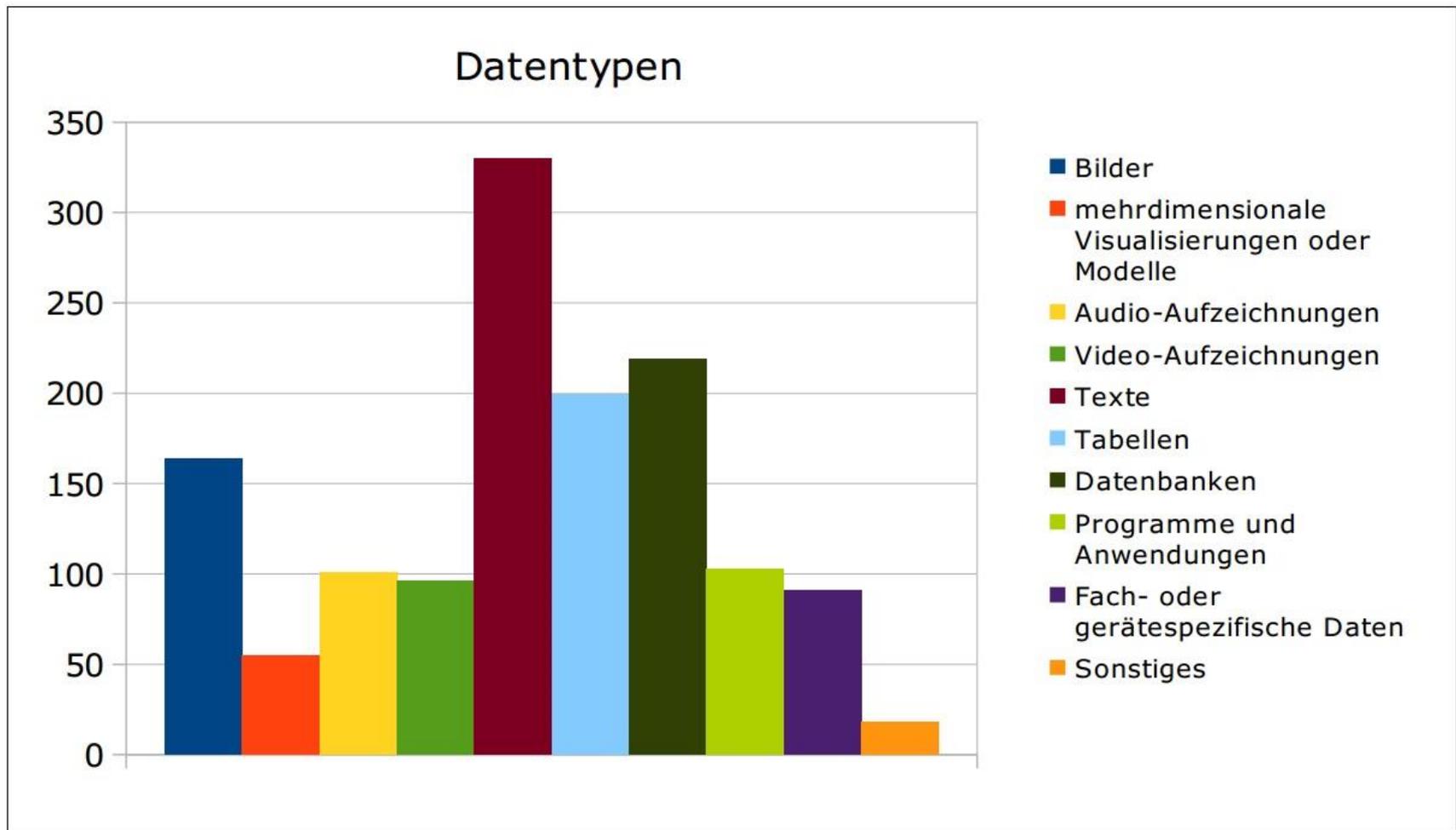


Figure 15: data types used by researchers, n = 1366

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



*Abbildung 4: Typen von Forschungsdaten*

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

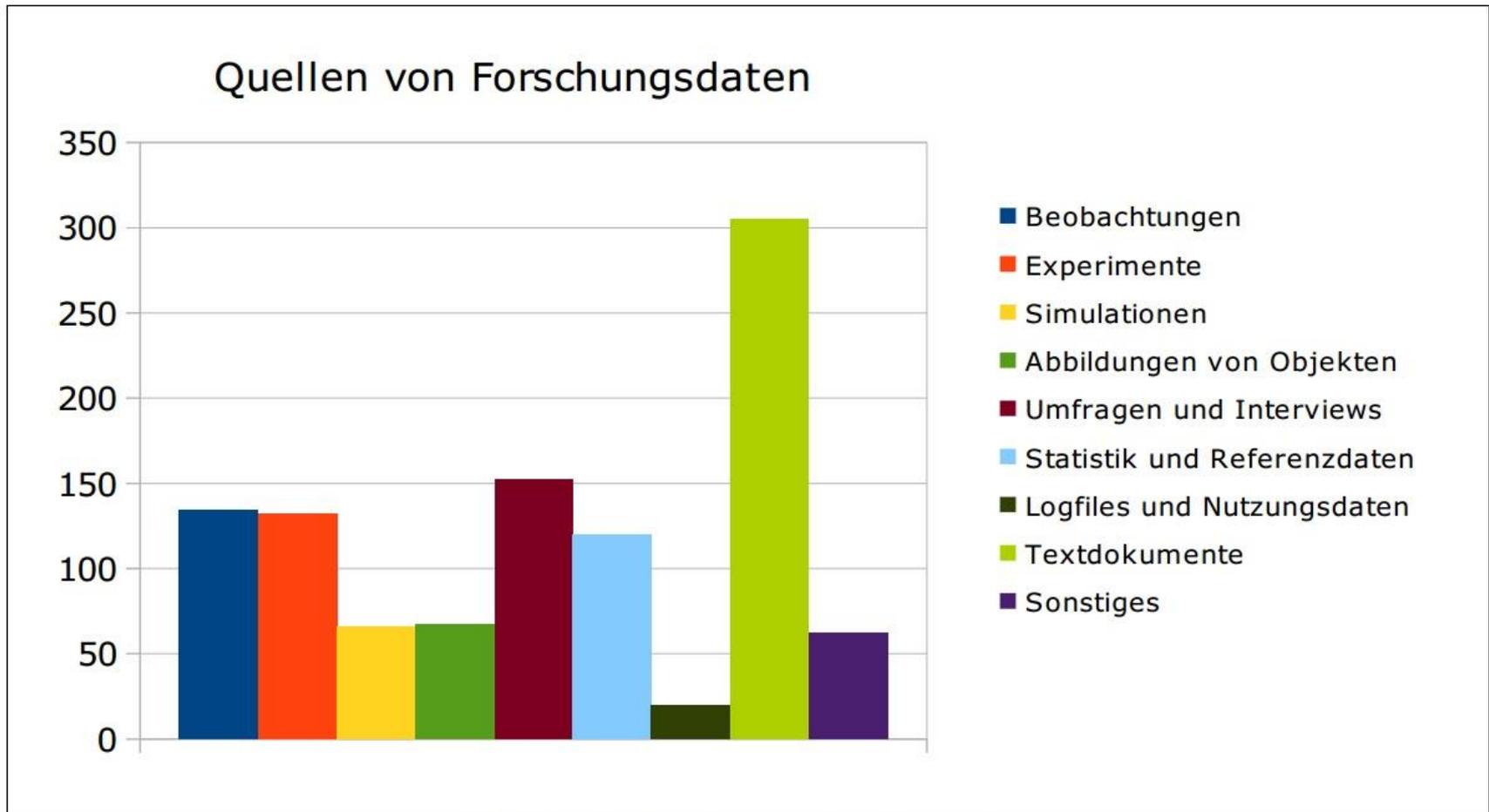
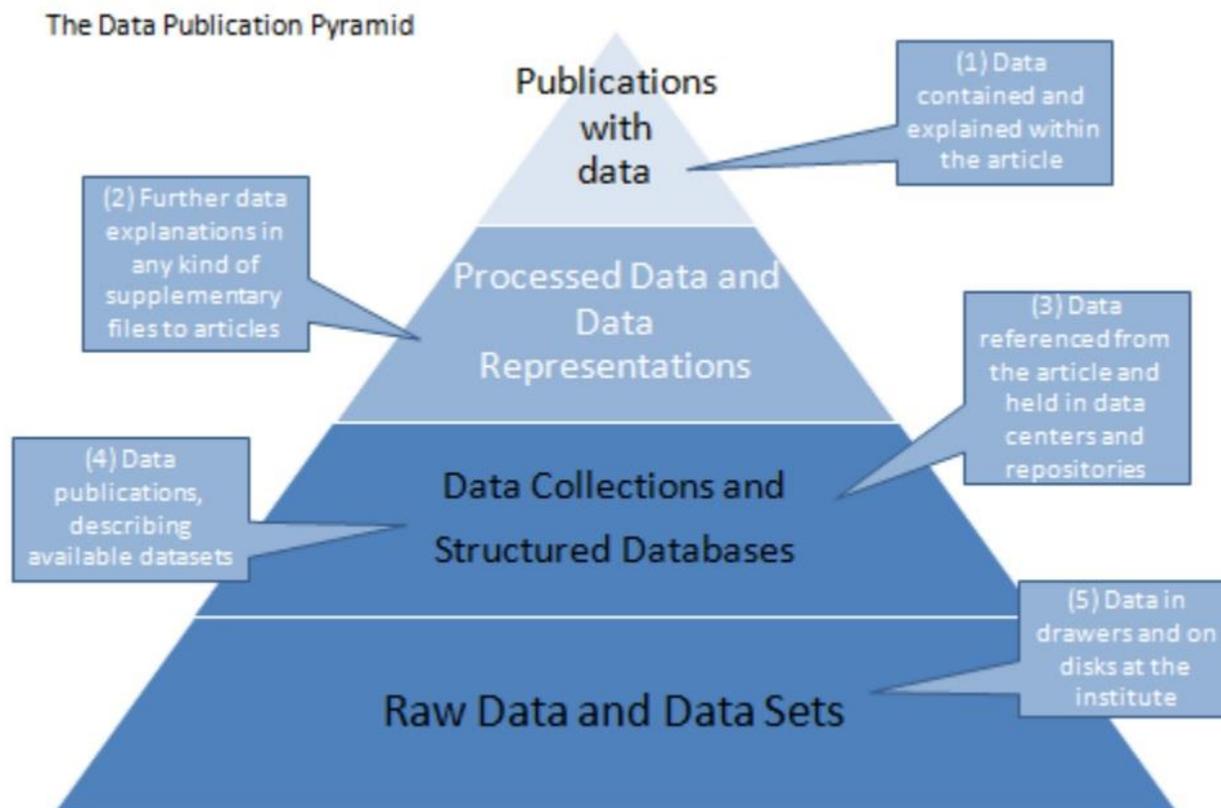


Abbildung 3: Quellen von Forschungsdaten

# 3. Spielarten und Definition von Forschungsdaten

*Report on Integration of Data and Publications*



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

### 3. Spielarten und Definition von Forschungsdaten



### 3. Spielarten und Definition von Forschungsdaten

*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.*

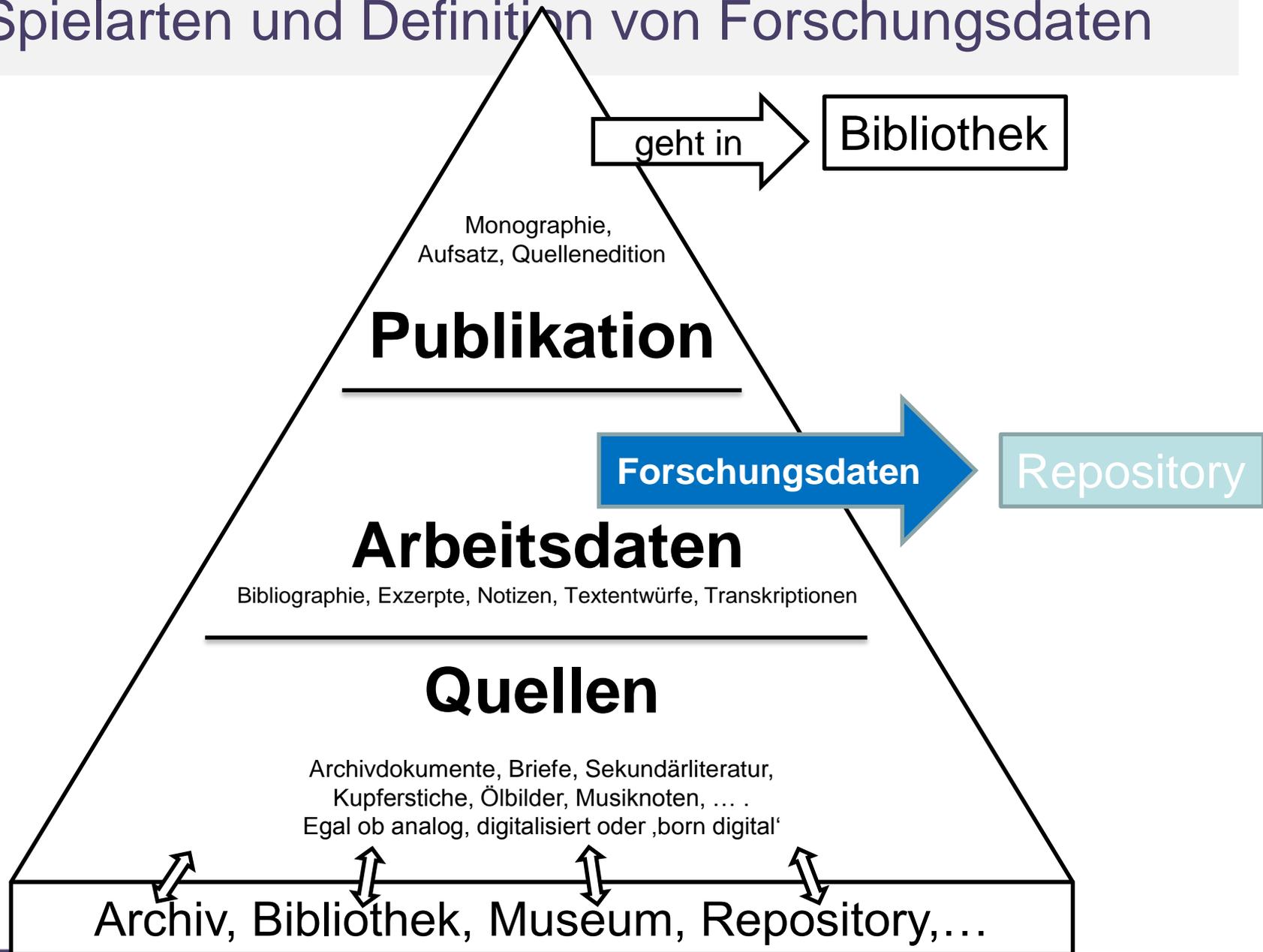
# 3. Spielarten und Definition von Forschungsdaten



### 3. Spielarten und **Definition** von Forschungsdaten

*Geisteswissenschaftlich Forschungsdaten sind alle jene Dateien, 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.*

### 3. Spielarten und Definition von Forschungsdaten



# 4. Ein Beispiel und eine Vision

The screenshot shows a Windows File Explorer window titled "Forschungsdaten". The address bar indicates the path: "Dieser PC > Lokaler Datenträger (C:) > Benutzer > csae8092 > Dropbox > Dokumente\_D > dariah > papers > Forschungsdaten". The left sidebar shows the navigation pane with "Lokaler Datenträger (C:)" selected. The main area displays a list of 38 PDF files, including titles like "Lauer, Geisteswissenschaften als Digital Humanities, 2013.pdf" and "Sudamih, FinalReport\_v1.0, 2011.pdf". The status bar at the bottom shows "38 Elemente" and "76.8 MB".

38 Elemente  
38 Elemente (Freier Speicherplatz: 1,50 GB) 76.8 MB Computer

# 4. Ein Beispiel und eine Vision

Navigation

Dokument durchsuchen

ÜBERSCHRIFTEN | SEITEN | ERGEBNISSE

- Heiko Tjalsma/ Jeroen Rombouts, Selection of Res...
  - 2.1.1 Primary and secondary research data
- Susann Reilly u.a., Report on Integration of Data an...
  - 1.4 The Data Publication Pyramid:
    - University of Minnesota Libraries, A Multi-Dimensi...
      - PARSE-REPORT
    - Borgmann, Data Scholarship in the Humanities, 20...
- Hügi/Schneider, Digitale Forschungsinfrastrukture...
  - 3. Forschung
    - 3.1. Forschung in den Geisteswissenschaften
      - 3.3. Forschungsdaten
- Burrows, Sharing humanities data
- Patrick Sahle, Jenseits der Daten. 2012
- Julia Behrens, Lars Fischer, Karl-Heinz Minks, Die in...
  - 1.2 Zur Ein- und Abgrenzung der Geisteswissen...
- Wolfgang Pempe, Geisteswissenschaften, in: Heike...
- Gerhard Lauer, Die digitale Vermessung der Kultur...
  - [DH-Fragstellungen und Hermeutik]
- Alan Liu, The Meaning of the Digital Humanities, i...
- Rens Bod/Jaap Maat/Thijs Weststeijn, The Making...
- Anne Burdick u.a., Digital Humanities, Cambridge...
  - Humanities to Digital Humanities
- Aleida Aassmann, Einführung in die Kulturwissens...

Heiko Tjalsma/ Jeroen Rombouts, Selection of Research Data. Guidelines for appraising and selection research data (DANS Studies in Digital Archiving 6), Den Haag 2011.

online:

<http://www.dans.knaw.nl/nl/over/organisatie-beleid/publicaties/DANSselectionofresearchdata.pdf>

### 2.1.1 Primary and secondary research data

There is an essential difference between primary and secondary data.

**Primary data** are data in their most basic and elementary form: unembellished, pure observations. These are often **raw data**, i.e. **data not yet influenced or edited by researchers**. Once **researchers do something** to these primary data, **they become secondary data**. Secondary data, or **processed data**, can be a combination or recombination of data or data that are recoded, categorised or visualised. Secondary data are often the data that are communicated to the outside world in one way or another, depending on the discipline or domain, either in publications (enhanced or otherwise) or in data laboratories.<sup>1</sup>

**De Cock Buning, The legal status of research data, 2011**

online<sup>2</sup>

Research data

**andorfer**  
primary/raw data = Quellen  
processed/secondary data = Fors  
abzugrenzen von der Publikation

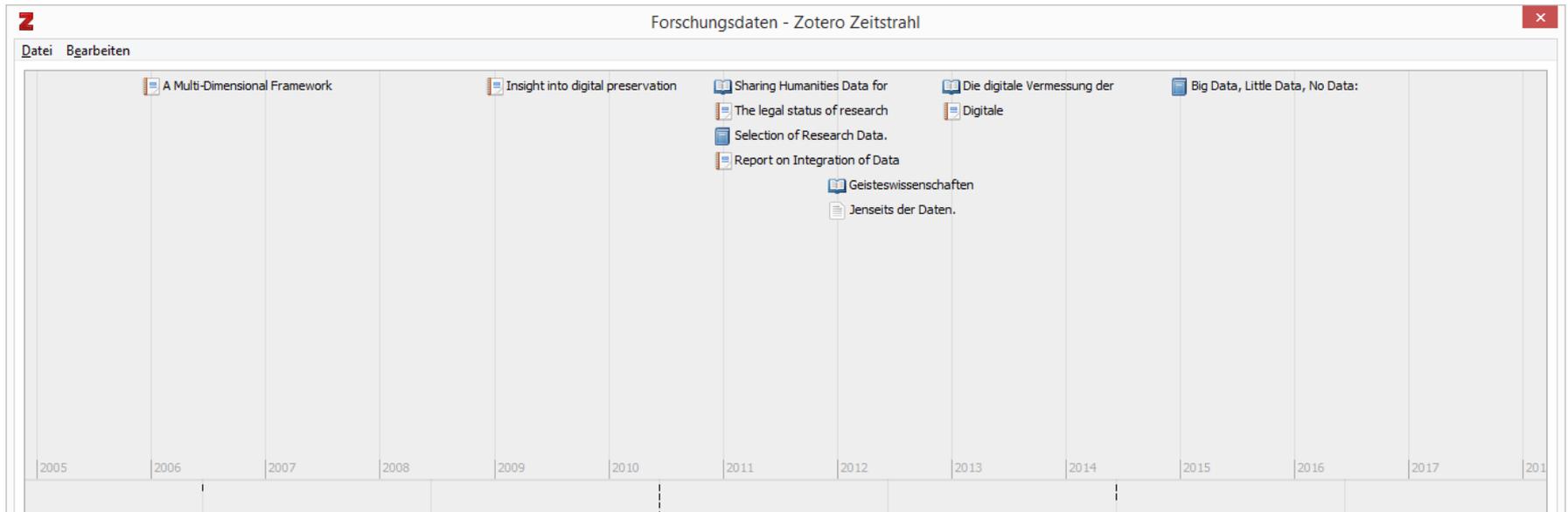
SEITE 1 VON 29 9660 WÖRTER ENGLISCH (USA) 110%

# 4. Ein Beispiel und eine Vision

The screenshot shows the Zotero web interface. At the top, the Zotero logo is on the left, and a navigation bar includes 'Home', 'My Library', 'Groups', 'People', 'Documentation', 'Forums', and 'Get Involved'. A search bar is on the right. Below the navigation bar, the breadcrumb path is 'Home > People > Peter Andorfer > Library > Forschungsdaten'. The left sidebar shows a library structure with 'DARIAH' and 'Forschungsdaten' selected. Below that is a 'Tags' section with a filter box and a grid of tag buttons. The main content area displays a table of items in the 'Forschungsdaten' library.

	Title	Creator	Item Type	Year	Date Added
<input type="checkbox"/>	Sharing Humanities Data for e-Research. Conceptual and Techn...	Burrows	Book Section	2011	2/24/2015 10:22 AM
<input type="checkbox"/>	Big Data, Little Data, No Data: Scholarship in the Networked...	Christine	Book	2015	2/24/2015 9:24 AM
<input type="checkbox"/>	The legal status of research data in the Knowledge Exch...	de Cock Buning	Report	2011	2/24/2015 9:10 AM
<input type="checkbox"/>	Digitale Forschungsinfrastrukturen für die Geistes- und Gesc...	Hügi and Schneider	Report	2013	2/24/2015 10:19 AM
<input type="checkbox"/>	Insight into digital preservation of research output in Euro...	Kuipers and van der Hoeven	Report	2009	2/24/2015 9:18 AM
<input type="checkbox"/>	Die digitale Vermessung der Kultur. Geisteswissenschaften al...	Lauer	Book Section	2013	2/24/2015 10:36 AM
<input type="checkbox"/>	The Meaning of the Digital Humanities	Liu	Journal Article		2/24/2015 10:40 AM
<input type="checkbox"/>	Geisteswissenschaften	Pempe	Book Section	2012	2/24/2015 10:33 AM
<input type="checkbox"/>	Report on Integration of Data and Publications	Reilly	Report	2011	2/24/2015 9:13 AM
<input type="checkbox"/>	Jenseits der Daten. Überlegungen zu Datenzentren für die Gei...	Sahle and Kronenwett	Journal Article	2012	2/24/2015 10:26 AM

# 4. Ein Beispiel und eine Vision



# 4. Ein Beispiel und eine Vision

Behrens, internationale  
Positionierung der  
Geisteswissenschaften  
2010

„Grenze zwischen GW  
und NW verschwimmt“

wird zitiert

Pempe,  
Geisteswissenschaften  
2012

wird zitiert

Hügi,  
Forschungsinfrastruktur  
2013

wird zitiert

Sahle, Jenseits der  
Daten,  
2013