

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

---

---

2016

### Big Data in den Geisteswissenschaften: Konzept für eine Lehr- und Lernmittelsammlung

Mareike Schumacher (Hamburg), Marcus Held (Mainz), Claudia Falk (Mainz), Stefan  
Pernes (Würzburg)

DARIAH-DE Working Papers

Nr. 15

Schumacher, Mareike; Held, Marcus; Falk, Claudia; Pernes, Stefan: Big Data in den Geisteswissenschaften: Konzept für eine Lehr- und Lernmittelsammlung  
Göttingen : GOEDOC, Dokumenten- und Publikationsserver der Georg-August-Universität, 2016  
(DARIAH-DE working papers 15)

Verfügbar:

PURL: <http://resolver.sub.uni-goettingen.de/purl/?dariah-2016-1>

URN: <http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2016-1-2>

Dieser Beitrag erscheint unter der Lizenz [Creative-Commons Attribution 4.0 \(CC-BY\)](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:** Dieses Working Paper beschreibt die Inhalte, Darstellungsformen und die medialen Umsetzungsmöglichkeiten einer Lehrmittelsammlung zum Thema „Big Data Methodik in den Geistes- und Kulturwissenschaften“. Zudem wird ein Disseminationskonzept entwickelt, das aufzeigt, auf welche Weise Inhalte, Themen und Instrumente dieses transdisziplinären Bereichs in den jeweiligen Fachdisziplinen vermittelt werden können. Das hier vorgestellte Konzept einer Lehr- und Lernmittelsammlung ist auf die Nutzung und Anwendung von Big Data-Technologien und Methoden für geistes- und kulturwissenschaftliche Forschungsfragen ausgerichtet.

**Keywords:** Big Data, Digitale Geisteswissenschaften, Geisteswissenschaften, Methodenvermittlung  
Big data, digital humanities, humanities, dissemination of methods

# Big Data in den Geisteswissenschaften: Konzept für eine Lehr- und Lernmittelsammlung

Mareike Schumacher (Hamburg), Marcus Held  
(Mainz), Claudia Falk (Mainz), Stefan Pernes  
(Würzburg)



Mareike Schumacher, Marcus Held, Claudia Falk, Stefan Pernes: „Big Data in den Geisteswissenschaften: Konzept für eine Lehr- und Lernmittelsammlung“. *DARIAH-DE Working Papers* Nr. 15. Göttingen: DARIAH-DE, 2016. URN: nbn:de:gbv:7-dariah-2016-1-2.

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**

Dieses Working Paper beschreibt die Inhalte, Darstellungsformen und die medialen Umsetzungsmöglichkeiten einer Lehrmittelsammlung zum Thema „Big Data Methodik in den Geistes- und Kulturwissenschaften“. Zudem wird ein Disseminationskonzept entwickelt, das aufzeigt, auf welche Weise Inhalte, Themen und Instrumente dieses transdisziplinären Bereichs in den jeweiligen Fachdisziplinen vermittelt werden können. Das hier vorgestellte Konzept einer Lehr- und Lernmittelsammlung ist auf die Nutzung und Anwendung von Big Data Technologien und Methoden für geistes- und kulturwissenschaftliche Forschungsfragen ausgerichtet.

# Inhaltsverzeichnis

1. Einleitung.....	4
2. Vom DARIAH-DE Helpdesk zur projektspezifischen Beratung.....	6
3. Inhalte der Lehrmittelsammlung.....	9
3.1 Anwendungsbeispiele (Use Cases).....	9
3.2 DARIAH-DE Schulungsmaterialien.....	11
3.3 Lehr- und Lernmaterialien als Open Educational Resources (OERs).....	13
4. Darstellungsformen der Lehrmaterialsammlung.....	14
5. Mediale Umsetzung der Lehrmaterialsammlung.....	18
6. Ziele und Herausforderungen.....	20
6.1 Zielformulierung.....	20
6.2 Qualitätssicherung.....	20
6.3 Datenbank- und Sammlungsarchitektur.....	21
7. Literatur- und Webseitenverzeichnis.....	22

# 1. Einleitung

Im Gegensatz zum herkömmlichen Verständnis von „Big Data“, wird der Begriff in den Geistes- und Kulturwissenschaften weniger im Zusammenhang mit tatsächlich sehr großen Datenmengen verwendet als vielmehr mit bestimmten datenzentrierten Methoden und der Verwendung entsprechender Tools. So meint der Begriff z.B. in den Sprach- und Literaturwissenschaften eine Abkehr von Einzelfallstudien sowie Close Reading und eine Hinwendung zu Korpusanalysen und Distant Reading (vgl. Moretti 2013). In den Bildwissenschaften dienen diese Methoden und Tools etwa dazu, bis zu mehrere Tausende Bilder gleichzeitig zu visualisieren und zu untersuchen. In den Geschichtswissenschaften lassen sich Datenbanken wie Wikimedia nach bestimmten biografischen Daten durchsuchen und soziale Trends, wie z.B. die Entwicklung von Berufsbildern oder auch Lebenszeiterwartungen, quantitativ auswerten. Im Folgenden nutzen wir die Begriffe „Big Data Methodik“ oder aber „große Datenmengen“, um sowohl die Anlehnung an die impulsgebenden Trends beizubehalten als auch eine gewisse Nähe zur tatsächlichen Datenbasis von Forschungsprojekten in den Geistes- und Kulturwissenschaften zu wahren.

In diesem Working Paper stellen wir neben den Inhalten, Darstellungsformen und den medialen Umsetzungsmöglichkeiten einer Lehrmittelsammlung zum Thema „Big Data Methodik in den Geistes- und Kulturwissenschaften“, ein Disseminationskonzept vor, das die Vermittlung von Inhalten, Themen und Instrumenten dieses transdisziplinären Bereichs in den jeweiligen Fachdisziplinen aufgreift. Das vorgestellte Konzept einer Lehr- und Lernmittelsammlung hat prototypisch-generischen Charakter und ist auf die Nutzung und Anwendung von Big Data Technologien und Methoden für geistes- und kulturwissenschaftliche Forschungsfragen ausgerichtet. Dabei berücksichtigen wir, dass sich Inhalte und Fragestellungen von Forschungsprozessen durch die Anwendung von Big Data Methodik und den Einsatz datenbasierter Tools verändern. Zugleich wird bedacht, dass die Beschaffenheit der Daten in ihrer Vielfalt und Komplexität variieren und bestimmten fachbereichsspezifischen Entwicklungen unterliegen kann. Angesichts der dadurch erhöhten Anforderungen an Didaktik, Vermittlungsmethoden und -techniken sowie der Variabilität und Komplexität der Daten werden zwei prototypisch-generische Strategien entwickelt:

1. Die Erprobung quantitativer Analysemethoden anhand unterschiedlicher disziplinärer Perspektiven und Anforderungen. Als Beispiele dienen uns die Anwendungsfälle (Use Cases), die im Rahmen des Forschungs- und Infrastrukturprojekts DARIAH-DE II<sup>1</sup> umgesetzt werden.

---

<sup>1</sup> DARIAH-DE (Digital Research Infrastructure for the Arts and Humanities) bezeichnet das deutsche Teilprojekt innerhalb von DARIAH-EU, das digital arbeitende Geistes- und KulturwissenschaftlerInnen mit Werkzeugen der Digital Humanities (DH) und fachwissenschaftlichen Diensten in den Bereichen Forschung und Lehre unterstützt. Das Projekt erzeugt u.a. eine digitale Forschungsinfrastruktur für Tools und Forschungsdaten und stellt Materialien für Lehre und Weiterbildung im Bereich der Digital Humanities bereit (<https://de.dariah.eu/>).

2. Die bisherigen Instrumente, Methoden und Dienste der Vermittlung von Methodenkompetenz, die bereits durch DARIAH-DE bereitgestellt wurden, werden analysiert und ergänzt.

Das Konzept der Lehr- und Lernmittelsammlung ist aus einer nutzerzentrierten Perspektive heraus entwickelt worden und stellt eine Grundlage für die Erstellung einer Lehrmittelsammlung dar. Als NutzerInnen sind vornehmlich Geistes- und KulturwissenschaftlerInnen gemeint, die in ihren Projekten bereits mit großen Datensammlungen arbeiten oder zukünftig arbeiten möchten.

Der hier entwickelte Ansatz entspringt dem Versuch, möglichst viele Fachbereiche und Anforderungsniveaus zumindest implizit zu berücksichtigen. So soll die Akzeptanz in den jeweiligen geistes- und kulturwissenschaftlichen Fachcommunitys erhöht und die Verbreitung durch prototypische Angebote in Form von vorgelegten Fallstudien und angebotenen Veranstaltungen unterstützt werden. Diese Angebote sollen sowohl den Einstieg in die Nutzung digitaler Tools und Methoden erleichtern als auch denjenigen NutzerInnen zur Weiterbildung dienen, die bereits mit Methoden der Digital Humanities (DH) arbeiten. Außerdem ist vorgesehen, den NutzerInnen ein Forum für den Austausch von Ideen anzubieten und auf diese Weise methodologisch neue Impulse zu setzen. Dies kann durch die Organisation von Veranstaltungen sowie durch die Beteiligung einzelner ForscherInnen an der Erstellung des vorgesehenen Lehr- und Lernbuches verwirklicht werden.

Dieses Lehrbuch ist als intern verlinkte Sammlung von Tutorials zu Daten, Methoden und Tools konzipiert und basiert auf den Lehrmitteln des DARIAH-DE-Portals (<https://de.dariah.eu/schulungsmaterial>), die die Big Data Methodik fokussieren und referenzieren. Zur Taxonomisierung der anzulegenden Lehr- und Lernmittelsammlung soll die Taxonomy of Digital Research Activities in the Humanities (TaDiRAH)<sup>2</sup> (<http://tadirah.dariah.eu/vocab/>) verwendet werden. Durch die Nutzung von TaDiRAH ist es außerdem möglich, Annotationen vorzunehmen und damit die Auszeichnung von Tool-Verzeichnissen über die Anreicherung bibliografischer Angaben mit der gleichzeitigen Reflexion der digitalen Forschungspraxis im Allgemeinen und im Hinblick auf das projektspezifische Anforderungsprofil im Besonderen zu verbinden. Auf diese Weise lassen sich bereits existierende, aber oft isoliert vorliegende Lehr- und Lernmittel sowohl generisch als auch projektspezifisch durchsuchen, vergleichen und nutzen. Das Lehr- und Lernbuch stellt somit auch spezifische Projektanwendungsfälle vor.<sup>3</sup>

Indem das Disseminationskonzept mit dem DARIAH-DE Helpdesk verknüpft wird, lässt sich die projektspezifische Beratung einbinden. Damit können die Aktivitäten von DARIAH-DE als Teil einer übergreifenden Infrastruktur in den Geistes- und Kulturwissenschaften etabliert werden.

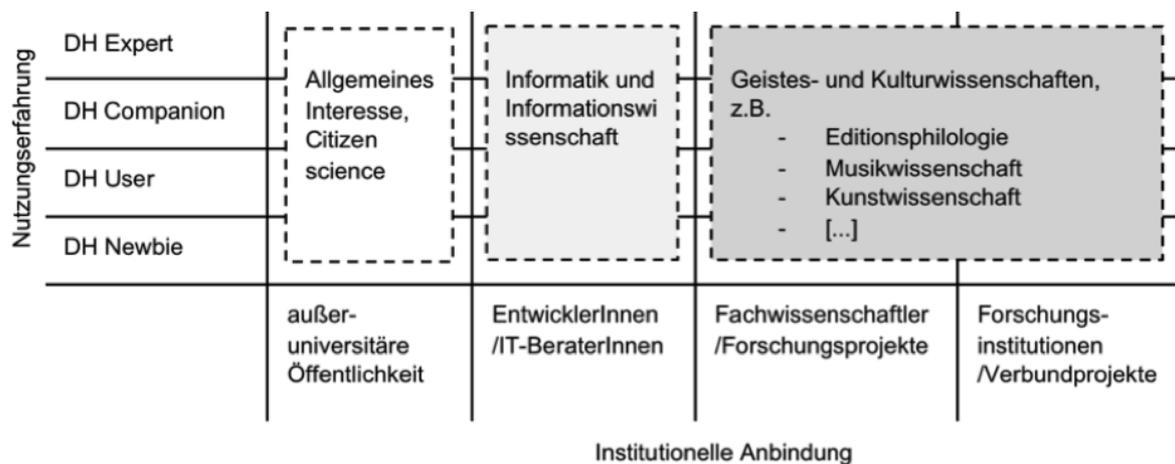
---

2 Für einen Überblick zur Entstehung und Entwicklung von TaDiRAH siehe Borek, Luise u.a.: „TaDiRAH: a case Study in Pragmatic Classification“. In: Digital Humanities Quarterly 10.1, 2016. Online: <http://digitalhumanities.org/dhq/vol/10/1/000235/000235.html>.

3 Ein ähnliches Konzept der medialen Doppelläufigkeit verfolgte auch das Projekt von L3T: Lehrbuch für Lernen und Lehren mit Technologien (<http://l3t.eu/homepage>).

## 2. Vom DARIAH-DE Helpdesk zur projektspezifischen Beratung

Im Rahmen der DARIAH-DE e-Humanities Infrastructure Service Unit (DeISU) wird ein Helpdesk etabliert. Die Hauptaufgabe des Helpdesks soll darin bestehen, je nach Zuständigkeitsbereich persönliche Berater zu vermitteln.



**Abbildung 1: Nutzergruppen.**

Die persönliche Beratung kann je nach NutzerInnen auf zwei Ebenen stattfinden:

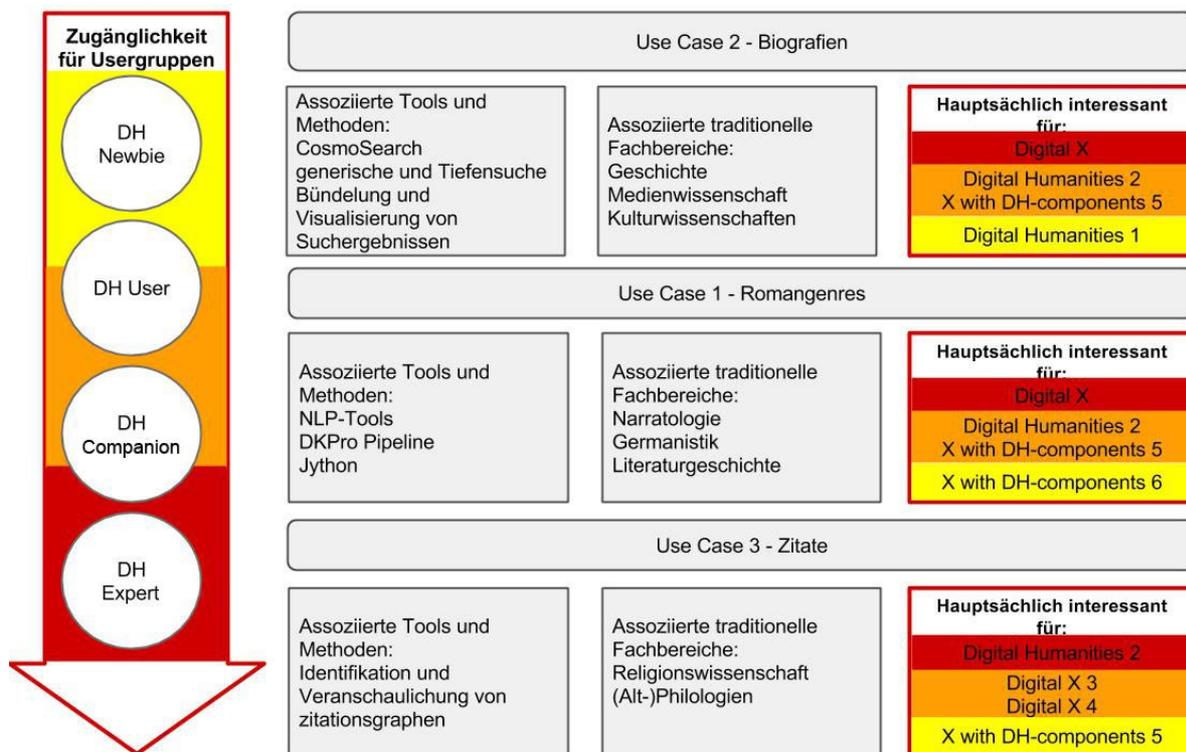
1. Weitere persönliche Beratung,
2. Hilfe zur Selbsthilfe durch Verweise auf zwei bis drei relevante Lehr- und Lernmittel mittels der TaDiRAH-Taxonomie.

Das Beratungsrepertoire muss von einem Low-Level-Entry ausgehen und ist dann nach Nutzergruppen steigerbar; dabei folgt das DARIAH-DE Helpdesk den Abläufen, die in der DH Mastermind Mind Map dargestellt sind (<https://www.mindmeister.com/de/137418763/dh-projects-public-version-1>).

Wie Patrick Sahle in seinem Report "DH studieren!" (vgl. Sahle 2013) deutlich macht, ist von unterschiedlichen Ausbildungsprofilen auszugehen, die bei dem Aufbau einer Lehr- und Lernmittelsammlung besonders berücksichtigt werden müssen. Ein niederschwelliger Einstieg soll möglich sein, um auch breitere Zielgruppen, die bisher noch nicht mit Verfahren, Methoden und Instrumenten der Digitalen Geisteswissenschaften vertraut sind, an die Materie heranzuführen. Um diesen Ansprüchen gerecht zu werden, soll die Lehr- und Lernmittelsammlung Materialien möglichst so aufbereiten, dass unterschiedliche Grade der Nutzung von DH-Methoden und Instrumenten Berücksichtigung finden. Die Gestaltung bzw. Aufbereitung der Materialien muss zudem verschiedene Zielgruppen ansprechen. Zu diesen Zielgruppen zählen FachwissenschaftlerInnen, die DH-Verfahren entwickeln und/oder anwenden, fachwissenschaftliche sowie „digitale GeisteswissenschaftlerInnen“ (digitale FachwissenschaftlerInnen, „Digital Humanists“) und auch FachwissenschaftlerInnen ohne

Erfahrungen in DH (vgl. Sahle 2013, 6).<sup>4</sup> Auch AbsolventInnen von DH-Programmen, die Sahle (vgl. 2013, 10f.) in sechs verschiedene Studiengruppen unterteilt, werden voraussichtlich verstärkt auf die Inhalte zugreifen und sicherlich mit sehr unterschiedlichen Erwartungen und Bedürfnissen eine Lehr- und Lernmittelsammlung zu Big Data Methoden nutzen.

Im Rahmen des Forschungsprojekts DARIAH-DE werden derzeit drei Anwendungsszenarien (Use Cases) erprobt, die allesamt in textbasierten Fachdisziplinen verortet sind (zu den Inhalten der Use Cases s.u.: „3 Inhalte der Lernmittelsammlung“). Die folgende schematische Darstellung zeigt, dass diese Anwendungen eine große Bandbreite an allgemeinen Nutzerinteressen abdecken. Sowohl die oben entwickelten vier Nutzergruppen, die den Grad der Kompetenz der DH Methoden in den Blick nehmen als auch die sechs Studiengruppen, die Patrick Sahle (vgl. 2013, 10f.) definiert, werden von den Use Cases (jeweils in unterschiedlicher Intensität) berücksichtigt:



**Abbildung 2: Use Cases in Bezug auf Nutzergruppen und DH Studienrichtungen.**

Der Versuch, diese anwendungszentrierte Darstellung mit der Idee eines DARIAH-DE Helpdesk zu verbinden, führt allerdings schnell zu der Einsicht, dass eine Entscheidung bezüglich der Ausrichtung auf Nutzerinteressen gefunden werden muss. Die derzeitige Arbeit an den Lehr- und Lernmitteln, die im Rahmen der Anwendungsszenarien eingesetzt werden, zielt vor allem darauf ab, einen Zustand zu erreichen, der der Darstellung in Abb. 2 entspricht. Durch die Homogenität der Use Cases wird auch eine gewisse Bandbreite an NutzerInnen erreicht. Das

<sup>4</sup> Die Unterscheidung orientiert sich an der von Patrick Sahle entwickelten Heuristik von DH-Studierenden, die allerdings auch das Gesamtfeld der bisherigen Akteursszenarien treffend nachzeichnet.

Konzept des Helpdesks basiert jedoch auf einem anderen Verständnis von Nutzerorientierung. So können NutzerInnen an das Helpdesk herantreten, die mindestens zwei sich überlappende Interessen haben – ein fachliches Interesse und eines, das die Ausprägung der eigenen Methodenkompetenz betrifft. Handelt es sich z.B. um LiteraturwissenschaftlerInnen, die noch neu innerhalb der DH sind und gerne 'Natural Language Processing' (NLP) Tools auf ein Korpus aus Erzähltexten anwenden möchten, müssten diese entweder eine ausführliche persönliche Anleitung bekommen oder ein extra auf ihren Wissensstand abgestimmtes Lernmittel, da das Anwendungsbeispiel eher für fortgeschrittene NutzerInnen aufbereitet ist. Für die zu entwickelnden Lernmittel bedeutet dies, dass sie in unterschiedlicher Aufbereitung vorliegen müssten, um alle vier Nutzergruppen bedienen zu können.

In der Praxis würde entsprechend zuerst durch das Helpdesk ein enger persönlicher Kontakt zu interessierten ForscherInnen hergestellt. Im Zuge der persönlichen Beratung sollte erfragt werden, ob das Interesse sich eher auf die eigene Forschung oder auf die Lehre bezieht. Ist ersteres der Fall, so muss als nächstes ermittelt werden, welches Vorwissen die Anfragenden haben. Im Falle von ForscherInnen die sich für Tools und Methoden des Natural Language Processing (NLP) interessieren, beträfe dies ein Grundverständnis von Statistik und mathematischen Prozessen sowie von maschinellem Lernen. Hinzu kämen eventuell Vorkenntnisse in einer Programmiersprache. Je nach Kenntnisstand der Anfragenden kann auf Lernmittel der eigenen Sammlung sowie ggf. auf Lektürearregungen aus den entsprechenden Bereichen verwiesen werden.

Sind ForscherInnen nicht daran interessiert, die eigene Forschungsperspektive zu erweitern, sondern suchen nach Anregungen zur Vermittlung von Tools und Methoden der Digital Humanities, benötigen sie eine grundlegend andere Art der Beratung. Zwar ist auch hier zunächst zu erfragen, welche Kenntnisse im angefragten Gebiet vorhanden sind, allerdings ist eher davon auszugehen, dass nur ForscherInnen DH lehren möchten, die zumindest über ein Grundverständnis der angewandten Verfahren verfügen. Auch in diesem Fall sollte allerdings das Gebiet ermittelt werden, über das gelehrt werden soll. Als Lösungsoption bietet sich in einem solchen Fall eher selten eines der Lernmittel an, die im Rahmen von DARIAH-DE erprobt wurden. Stattdessen könnten den Anfragenden bestimmte DARIAH-Veranstaltungen empfohlen werden, die der Methodenkompetenzvermittlung dienen. Ist das Interesse breit gefächert, böte sich beispielsweise der Besuch der EDIROM Summerschool an, möchten sich die Anfragenden auf ein spezielles Gebiet konzentrieren, so wird ihnen der Besuch eines DARIAH-Methodenworkshops weiterhelfen.

### 3. Inhalte der Lehrmittelsammlung

Die Inhalte der Lehrmittelsammlung gliedern sich in Materialien, die 1. im Rahmen der genannten Anwendungsbeispiele erarbeitet werden und 2. bereits als Schulungsunterlagen in der DARIAH-DE-Sammlung von Schulungsmaterialien (<https://de.dariah.eu/schulungsmaterial-sammlung>) sowie in der Bibliographie „Doing Digital Humanities“ (<https://de.dariah.eu/bibliographie>) vorliegen. Im Folgenden stellen wir die einzelnen Bestandteile näher vor und beschreiben, welche Inhalte zur Schulung unterschiedlicher Nutzergruppen dienen können.

#### 3.1 Anwendungsbeispiele (Use Cases)

Ausgehend von einer Beschreibung der Anwendungsbeispiele, die im Rahmen von DARIAH-DE bearbeitet wurden, lassen sich Unterschiede und Gemeinsamkeiten der Use Cases im Hinblick auf Forschungsprozesse, -methoden und -objekte herausarbeiten. Im ersten Schritt werden die Ansätze der Anwendungsfälle als prototypische DH-Fragestellungen mit Big Data Methodik vorgestellt, im zweiten Schritt lassen sich dann übergreifende Inhalte identifizieren, um daraus die inhaltliche Struktur der Lehr- und Lernmittelsammlung abzuleiten.

Das Anwendungsbeispiel „Narrative Techniken und Untergattungen im deutschen Roman“ (Use Case 1) zeigt exemplarisch, wie eine große Sammlung literarischer Texte genutzt werden kann, um mit Hilfe quantitativer Verfahren die historische Entwicklung narrativer Techniken und in weiterer Folge auch die Entwicklung darauf aufbauender Kategorien zu analysieren. Zum Einsatz kommen Verfahren zur automatischen Erkennung bestimmter Merkmale, wie zum Beispiel Eigennamen oder Passagen direkter Rede. Die Merkmale werden im Anschluss zueinander in Bezug gesetzt, um die Texte zu gruppieren und narratologische Kategorien zu prüfen.

Ein Ziel dieser Anwendung ist es, die Entwicklung der Werkzeuge auch in Form von Lernmaterialien abzubilden, die verschiedene Einstiegspunkte bieten – etwa für AnwenderInnen mit weniger Erfahrung im Umgang mit Programmiersprachen. Die technische Umsetzung teilt sich in zwei Bereiche und entlastet die NutzerInnen: Sämtliche Schritte, die zur Verarbeitung von Natural Language Processing (NLP) nötig sind, werden mithilfe des DKPro Frameworks der TU Darmstadt als eine generische Pipeline, d.h. als Abfolge aufeinander aufbauender Verarbeitungsschritte, umgesetzt und in Form eines Plaintextformats ausgegeben. Dieses kann programmatisch einfach eingelesen und auch mit herkömmlicher Office Software geöffnet werden. Ab diesem Punkt beginnt die Weiterverarbeitung der Ergebnisse, im Use Case prototypisch umgesetzt anhand der Skriptsprachen Python und R.

Folgende Anforderungen ergeben sich für die BenutzerInnen: Die Werkzeuge des Anwendungsbeispiels und somit auch die Lernmittel, die daraus hervorgehen, konzentrieren sich auf Programmcode ohne grafische Benutzerführung. Grundlagen der Programmierung, der Statistik und des Natural Language Processing (NLP) sind demnach für den Einstieg hilfreich, jedoch nicht notwendig, da für alle relevanten Konzepte und Methoden Links zu weiterführenden Informationen bereitgestellt werden.

Im Rahmen des Anwendungsfalls „Biografien“ (Use Case 2) werden Verbindungen von individuellen historischen Lebensläufen und Internationalitätskriterien auf Grundlage von

Wikipedia und mehreren europäischen Nationalbiografien untersucht. Für dieses Beispiel, das eng mit dem Projekt „Cosmobilities – Grenzüberschreitende Lebensläufe in den europäischen Nationalbiographien des 19. Jahrhunderts“<sup>5</sup> verbunden ist, werden Frageraster und Kategorien für die Analyse serieller, digital verfügbarer Massendaten in der Biografieforschung entwickelt. Die technische Umsetzung des Use Case verbindet die Gewinnung von Informationen aus strukturierten und unstrukturierten Daten sowie aus mehrsprachigen Quellen. Die Analyse biographischer Daten mittels halbautomatischer Erkennung von Korrelationen zwischen Personen, Orten und Ereignissen, ermöglicht es schließlich, personenbezogene Bewegungsprofile zu erstellen und durch deren Aggregation spezifische Fragestellungen zu unterstützen. Zudem kann das System als wiederverwendbares Paket auch in anderen Kontexten eingesetzt werden, was allerdings eine Anpassung auf andere Datenbestände voraussetzt.

Folgende Anforderungen ergeben sich für die BenutzerInnen: Die Grundfunktionalität, den Bestand gewonnener Personenprofile samt ihrer Korrelationen durchsuchen zu können, ist über eine grafische Benutzeroberfläche zugänglich. Der Umgang mit dieser Suchfunktion wird in einer beiliegenden Projektdokumentation vermittelt und umfasst auch die Übermittlung einiger Schlüsselworte für komplexere Abfragen (eine „query language“). Ist es jedoch das Ziel der AnwenderInnen, das System auf andere Datensätze oder inhaltliche Domänen anzupassen, ergeben sich Anforderungen, die ähnlich gelagert sind wie im ersten Anwendungsbeispiel (Use Case 1), diese unter Umständen jedoch noch übersteigen.

Gegenstand des dritten Anwendungsfalls (Use Case 3) „Zitationen“ ist die Erschließung vollständiger diachroner Zitationsspuren in einer großen Sammlung griechischer und lateinischer Texte. Diese decken einen Zeitraum von der frühesten Kirchengeschichte bis in das 15. Jahrhundert ab. Ziel des gewählten Ansatzes ist es, über eine Rekonstruktion der intertextuellen Beziehungen in ihrer Gesamtheit, zu neuen Erkenntnissen über die Entwicklung des Christentums und des abendländischen Kulturraumes zu gelangen. Die technische Umsetzung verläuft schrittweise, ausgehend von einer Alignierung direkter Zitate in gleichsprachigen Texten, hin zur Erkennung umfassender Paraphrasierungen in sprachlich heterogenem Material. Die entwickelten Verfahren und die ihnen zugrunde liegende massiv parallelisierte Architektur werden zur Nachnutzung bereitgestellt.

Folgende Anforderungen ergeben sich für die BenutzerInnen: Ähnlich wie für den Anwendungsfall „Biografien“ gilt, dass eine Grundfunktionalität – hier ein systematisches Durchsuchen der Sammlung hinsichtlich ihrer intertextuellen Eigenschaften – über eine grafische Benutzeroberfläche zugänglich ist, die von einem Überblick des Gesamtkorpus' bis hin zur Ebene einzelner Wortformen reicht. Der Umgang mit grafischen Bedienelementen wird in einer beiliegenden Projektdokumentation vermittelt. Abgesehen davon wird eine Programmierschnittstelle (API) bereitgestellt, die einen möglichst vollständigen Zugang auf das System ermöglicht. Eine eigene Installation und Anpassung des Systems ist prinzipiell möglich, übersteigt aufgrund der Komplexität der Architektur jedoch den Rahmen der hier zu berücksichtigenden Lehr- und Lernmittel.

---

<sup>5</sup> Das Projekt „Cosmobilities“ ist am Leibniz Institut für Europäische Geschichte, Mainz, angesiedelt; online: [http://www.ieg-mainz.de/Forschungsprojekte-----\\_site.site..ls\\_dir\\_nav.17\\_f.69\\_likecms.html](http://www.ieg-mainz.de/Forschungsprojekte-----_site.site..ls_dir_nav.17_f.69_likecms.html).

## 3.2 DARIAH-DE Schulungsmaterialien

Um einen Überblick über bereits vorhandene Schulungsmaterialien, ihren inhaltlichen Aufbau und mögliche Pfadabhängigkeiten zu erhalten, wurde für diesen Beitrag eine explorative Inhaltsanalyse einschlägiger Titel durchgeführt. Grundlage bildeten sowohl die Sammlung von Schulungsmaterialien (<https://de.dariah.eu/schulungsmaterial-sammlung>), die im Rahmen von DARIAH-DE kompiliert wurde, als auch die Zotero-basierte Bibliografie „Doing Digital Humanities“ von DARIAH-DE (<https://de.dariah.eu/bibliographie>). Das Vorgehen war wie folgt: Alle relevanten Einträge – d.h. jene, die sich thematisch mit Data Science beschäftigen und als Schulungsmaterial geeignet sind – wurden gesichtet. Anschließend wurden die Einträge weiter nach Thema und Schwierigkeitsgrad gefiltert und bekannte Titel zuerst herausgegriffen. Schließlich wurden die Inhalte verglichen und induktiv Gruppen gebildet. Dieses bewusste Auswahlverfahren wurde so lange wiederholt, bis sich eine Stichprobe bestehend aus 16 Titeln und eine Aufteilung in drei Gruppen, die drei Anforderungsniveaus entspricht, gebildet hatte. Folgende Aufstellung beschreibt die in den Schulungsmaterialien vorgefundenen Inhalte:

1. Anforderungsniveau A:<sup>6</sup>
  - GUI-Bedienung (diverse), tagsets, query languages, rudimentäres Dateimanagement, Visualisierung – es handelt sich dabei um Anleitungen für spezifische Werkzeuge mit grafischer Benutzerführung und andere spezifische Einzelfertigkeiten.
2. Anforderungsniveau B:<sup>7</sup>
  - Grundlagen der Programmierung: arithmetics, string operations, data structures, file I/O, control structures, modularity, object orientation,
  - Grundlagen der Statistik: measures of central tendency, measures of variation, random variables, probability distributions, statistical modeling, hypothesis testing, analysis of variance, plotting,
  - Grundlagen des Natural Language Processing: boilerplate removal, tokenization, word frequencies, keywords-in-context.

---

6 Titel zu A: Jannidis, Fotis und Schöch, Christof: „DigiVoy: TextGrids Digitale Bibliothek mit Voyant entdecken“. Online: <https://de.dariah.eu/digivoy>; Sinclair, Stéfán und Rockwell, Geoffrey: „Voyant Tools Tutorial Screencasts“. Online: <https://www.youtube.com/playlist?list=PLDCADF35691404F54>; Graham, Shawn u.a.: „The Historian’s Macroscope“ - working title. Under contract with Imperial College Press. Open Draft Version. Autumn 2013. Online: <http://www.themacroscope.org>; Schulserver Hessen: „SQL-Tutorial“. Online: <http://www.oercommons.org/courses/sql-tutorial-2/view>; Knox, Doug: „Understanding Regular Expressions“, *The Programming Historian*. Online: <http://www.oercommons.org/courses/the-programming-historian-2-understanding-regular-expressions/view>. (Letzter Zugriff auf alle hier genannten Online-Ressourcen: 14.06.2016)

7 Titel zu B: Khan Academy: „Probability and statistics“. Online: <https://www.khanacademy.org/math/probability>; Haslwanter, Thomas: „An Introduction to Statistics using Python“. 2016. Basel: Springer International Publishing; Groß, Peter und Peters, Benjamin: „R Reader. Arbeiten mit dem Statistikprogramm R“. 2009. Online: <http://cran.r-project.org/doc/contrib/Grosz+Peters-R-Reader.pdf>; Turkle, William J. Und Crimble, Adam: „Introduction to Python“, *The Programming Historian*. Online: <http://programminghistorian.org/lessons/>; Hochschule der Medien Stuttgart: „Einführung in Python“. Online: <https://www.hdm-stuttgart.de/~maucher/Python/html/>. (Letzter Zugriff auf alle hier genannten Online-Ressourcen: 14.06.2016)

### 3. Anforderungsniveau C:<sup>8</sup>

- Machine Learning: linear algebra, linear regression, gradient descent, logistic regression, simple bayes classifier, k-means clustering, hierarchical clustering, regularization, distance metrics, evaluation metrics, dimensionality, reduction, anomaly detection, neural networks, support vector machine, gibbs sampling, latent dirichlet allocation,
- Natural Language Processing: document indexing, document scoring, www scraping, building networks from text.

Anhand der Aufstellung lassen sich Pfadabhängigkeiten erkennen, so zum Beispiel, dass Grundlagen aus Statistik und Programmierung notwendig sind, um Machine Learning-Verfahren zu nutzen. Ein Blick in die betreffenden Lehr- und Lernmittel führt zu folgenden Hypothesen hinsichtlich ihres inhaltlichen Aufbaus: In der ersten und in der dritten Gruppe (Anforderungsniveau A und C) findet eher eine Vermittlung ausgewählter Aspekte statt – zum Beispiel zugeschnitten auf spezifische Anwendungsfälle oder als cursorischer Überblick über mögliche Verfahren. In beiden Fällen sind es die vielfältigen Voraussetzungen – sowohl für EinsteigerInnen als auch für fortgeschrittene AnwenderInnen –, die eine solche Struktur bedingen. Demgegenüber ist in der zweiten Gruppe (Anforderungsniveau B), deren Inhalte relativ klar abgegrenzte Grundlagen vermitteln, am ehesten ein lückenloser Aufbau vorzufinden.

Ausgehend von der hier untersuchten Stichprobe lässt sich feststellen, dass es – insbesondere für höhere Anforderungsniveaus – Best Practice ist, weiterführende Hinweise auf sämtliche Inhalte anzubieten, die für deren Verständnis notwendig sind. Auch Nebenaspekte, wie die Einrichtung der Arbeitsumgebung oder alternativer Programmpakete, müssen berücksichtigt werden. Erstellt man eine Anwendungsdokumentation sollte entsprechend ein „roter Faden“ definiert werden, der auf einer fachwissenschaftlichen Forschungsfrage basiert. Anschließend lassen sich Haupt- und Nebenaspekte unterscheiden, die in unterschiedlicher inhaltlicher Tiefe dargestellt werden können.

Neben der Entwicklung eigener Lehr- und Lernmittel in den Anwendungsbeispielen, gilt es, auch das bereits im Rahmen des Forschungsprojekts DARIAH-DE I erstellte und gesammelte Material mit Blick auf das Thema Data Science zu sichten, zu evaluieren und zu integrieren. Indem vorhandene Materialien evaluiert und integriert werden, soll schließlich eine eigene Bibliografie zu „Big Data Methoden in den Humanities“ entstehen, die eine Rubrik der Bibliografie „Doing Digital Humanities“ darstellen wird: Die bereits auf dem DARIAH-DE-Portal vorhandenen Materialien werden gesichtet, selektiert und kommentiert. Zudem werden sie mit Annotationen versehen und Kategorien zugeordnet, so dass ein umfassendes Set von Materialien

---

<sup>8</sup> Titel zu C: Ng, Andrew: „Machine Learning“, *Coursera*. Online: <https://class.coursera.org/ml-003/lecture>; Hochschule der Medien Stuttgart: „Data Mining mit Scikit Learn“. Online: <https://www.hdm-stuttgart.de/~maucher/Phyton/SklearnIntro/html/dataminingSklearn.html>; Riddell, Allen: „TAToM: Text Analysis with Topic Models for the Humanities and Social Sciences“. Online: <http://www.oercommons.org/courses/tatom-text-analysis-with-topic-models-for-the-humanities-and-social-sciences/view>; Karsdorp, Folgert u.a.: „Python Programming for the Humanities“. Online: <https://github.com/sonofmun/FK-python-course/blob/master/README.md>; Layton: „Text mining online data with scikit-learn“. Online: [https://www.youtube.com/watch?v=2AEM89ltcgY&feature=youtube\\_gdata](https://www.youtube.com/watch?v=2AEM89ltcgY&feature=youtube_gdata). (Letzter Zugriff auf alle hier genannten Online-Ressourcen: 14.06.2016)

bereitgestellt werden kann. Ergänzend lässt sich die Bibliografie „Doing Digital Humanities“ mit Inhalten und Materialien verbinden, die im Rahmen der Forschungsprojekte DARIAH-DE I und DARIAH-DE II erarbeitet und gesammelt wurden. Somit werden die Use Cases in einen Gesamtzusammenhang gebracht und es entsteht eine Bibliografie zu „Big Data Methoden“, die ihrerseits einen Beitrag zur Selbsthilfe darstellt. Für die Arbeit des Helpdesks wird dies von zentraler Bedeutung sein.

### **3.3 Lehr- und Lernmaterialien als Open Educational Resources (OERs)**

Die bisherige Sammlung soll zudem themenspezifisch in die aufzubauende Daten- und Sammlungsarchitektur integriert werden. Damit bildet sie selbst einen Anwendungsfall für die Integration und Evaluierung von Lehr- und Lernmaterialien als Open Educational Resources (OERs). Mit Blick auf die vier Ressourcen der OER (re-use, re-vise, re-mix und re-distribute), bilden die vorhandenen Daten und Materialien der Use Cases auch hinsichtlich der Anreicherung mit Metadaten und Standards, der Bewertung der Qualität und der Klärung von lizenzrechtlichen Fragen einen wichtigen Anwendungsfall. Durch Kategorisierung und Taxonomisierung lassen sich die vorhandenen Wissensressourcen zusammenführen und durch ein integratives Komponentensystem in der Datenbank- und Sammlungsarchitektur weiterentwickeln. Hierbei wird die vorhandene Taxonomie TaDiRAH (<https://github.com/dhtaxonomy/TaDiRAH>) eingesetzt und ausgebaut, so dass sowohl ein induktiver, explorativer als auch ein deduktiver, sehr zielgerichteter Gebrauch möglich sein wird. Damit verfolgt die Lehr- und Lernmittelsammlung einen individuellen, aber ganzheitlich angelegten Lehr- und Lernauftrag, der es ermöglicht, eine relativ breite Anzahl von Nutzergruppen abzudecken.

Das übergeordnete Ziel der Lehr- und Lernmittelsammlung ist es, eine geistes- und kulturwissenschaftliche Infrastruktur für OER weiterzuentwickeln. Im Fokus steht dabei die Analyse großer Datenmengen in unterschiedlichen disziplinären Kontexten und unter Berücksichtigung diverser fachwissenschaftlicher Perspektiven. Bei der Konzeption der Lehr- und Lernmaterialsammlung sowie der einzelnen Lehr- und Lernmittel sollte also berücksichtigt werden, dass das „technische Know How“ der jeweiligen Nutzergruppen sehr unterschiedlich ausfällt. Mit der doppelten Ausrichtung auf Vermittlung und Dissemination stellt die Sammlung ein Lehr- und Lern-System dar, das einerseits individuell, andererseits generisch, einerseits projektorientiert, andererseits modular angelegt ist.

## 4. Darstellungsformen der Lehrmaterialsammlung

Die Lehr- und Lernmittelsammlung wird als Infrastruktur des methodenorientierten Wissenstransfers und der Methodenkompetenzvermittlung verstanden. Innerhalb einer so konzipierten Sammlung sind besonders die Verknüpfungen einzelner Lehr- und Lernmittel wichtig. Die Sammlung wird so gestaltet, dass einerseits eine thematische Entwicklung von basalen zu spezifischen Kompetenzen möglich ist, andererseits NutzerInnen aber auch Querverbindungen zwischen thematisch verbundenen Einzelabschnitten folgen bzw. diese selbst aufgrund von spezifischen Aktivitätsmustern und Workflows herstellen können. Eine derartige Konzeption wurde beispielhaft für die bereits teilweise vorhandenen oder gerade in Planung befindlichen Lehr- und Lernmittel der Anwendungsbeispiele dargelegt; ergänzend kommen die Lehr- und Lernmaterialien zum Thema „Big Data Methoden in den Geisteswissenschaften“ hinzu, die bereits in der ersten Projektphase von DARIAH-DE erarbeitet wurden. Für den Aufbau und die Weiterentwicklung einer solchen Lehr- und Lernmittelsammlung wird vorgeschlagen, zweigliedrig vorzugehen:

1. Erstellung einzelner Lernmittel auf Grundlage der Anwendungsbeispiele,
2. Zusammenstellung und Herausgabe eines intermedial verfügbaren Lehr- und Lernbuches, welches von DARIAH-MitarbeiterInnen und VertreterInnen der Fachgesellschaften sowie ausgesuchten WissenschaftlerInnen erarbeitet wird und dann potentiell durch NutzerInnen ergänzt werden kann.

Beide Herangehensweisen folgen den Leitlinien von Digitalität, Anpassungsfähigkeit, Adaptierbarkeit und einer grundlegenden Offenheit. Durch diese Strategie ist es möglich, Wissen zu vermitteln und konkrete Lehr- und Lernprozesse zu unterstützen. Außerdem kann es auf diese Weise gelingen, in einem eigenen Forschungsprozess eine Sammlung aufzubauen und damit die Forschung an DH-Methoden, Tools und Materialien sowie deren Förderung voranzutreiben.

Die Lehr- und Lernmittel sollen in Form von Modulen bereitgestellt werden. Diese Modularisierung, die durch Annotieren und Kommentieren einzelner Komponenten und Module zusätzlich unterstützt werden kann, macht es möglich, verschiedenen Nutzerinteressen entgegenzukommen und bietet ein hohes Maß an Effizienz, Effektivität und Flexibilität. Gleichzeitig können die Nutzungsdaten für interne Analysen verwendet werden. In Folge einer systematischen Strukturierung ergeben sich so im besten Falle konkrete Lehr- und Lernszenarien, die aufzeigen, inwiefern einzelne Lehrmittel genutzt und (re-)konfiguriert werden, um sie den jeweiligen Anforderungen und Rahmenbedingungen der IndividualnutzerInnen anzupassen. Besonders interessant ist, inwiefern sich Materialien oder einzelne Bestandteile in verschiedenen Projektzusammenhängen wiederverwenden lassen, da dies ein konkreter Indikator für die Skalierbarkeit (vgl. Lehr 2011, 61ff.) der Nutzung ist. Um Modularisierungs- und Baukastenstrategien zu realisieren, müssen prinzipielle Standards und Regeln beachtet bzw. entwickelt werden. Hierbei spielen die Anreicherung mit Metadaten, die inhaltsbezogene Annotation und die Verwendung geeigneter Lizenzen eine wichtige Rolle. Diesbezüglich kann auf die Richtlinien zurückgegriffen werden, die innerhalb von DARIAH-DE im Rahmen der AG "Research Data Lifecycle" (<https://wiki.de.dariah.eu/x/cg9FAg>) ausgearbeitet wurden.

Ein Modul besteht aus mehreren Komponenten. Diese Aufteilung eines Moduls ermöglicht es, einzelne Lerninhalte aktuell zu halten, da Bestandteile einzeln angepasst werden können. Außerdem erhöht sich die Flexibilität der Nutzung; relevante Komponenten, wie z.B. Fallstudien, können ohne Weiteres einzeln verwendet werden, ohne dass dafür ein komplettes Modul angepasst werden müsste. Um die Vorteile solch einer Modularisierungsstrategie vollends nutzen zu können, muss diese übergreifend sowohl intern als auch extern in kooperativer als auch kollaborativer Hinsicht (weiter-)gedacht werden. Nur wenn Institutionen und Akteure innerhalb und außerhalb von DARIAH-DE in ständigem Austausch stehen, lassen sich Insellösungen vermeiden.

Es ist davon auszugehen, dass die bereitgestellten Lehr- und Lernmaterialien wesentlich interaktiver genutzt werden, wenn es möglich ist, Annotationen und Kommentare einzufügen. Auf diese Weise gelingt es, vom klassischen Lehrenden-Konzept abzuweichen und stattdessen die NutzerInnen als Lehrend-Lernende zu begreifen, die darüber hinaus daran mitwirken, die Inhalte der Lehr- und Lernmaterialsammlung in ein Linked Open Data (LOP) System zu überführen (vgl. Berners-Lee 2009). Dadurch werden die Inhalte miteinander zu einem sogenannten 'Liquid Content' verschmolzen. Die Lehr- und Lernmittelsammlung entspricht der derzeitigen Entwicklung hin zum Web 3.0, da semantische Netze für Inhalte, Methoden und Instrumente der Lehr- und Lernmittelsammlung genutzt und weiterentwickelt werden (vgl. Davies u.a. 2006; Fensel u.a. 2003; Daconta u.a. 2003).

Betrachtet man die Digitalisierung der Lehr- und Lernmittelsammlung mit ihren verschiedenen Komponenten als Liquid Content, ergeben sich für das Projekt inhaltlich zwei weitere Dimensionen:

1. Zunächst lassen sich Lehr- und Lernmittel inter- und multimedial nutzen. Außerdem wird eine interaktive Nutzung möglich: sowohl hinsichtlich der Arbeit mit Buch und Sammlung (z.B. durch Anlegen eines eigenen Portfolio) als auch bezüglich der Kooperations- und Kollaborationsmöglichkeiten (die Kommunikation kann dabei über Kommentare und Annotationen oder direkt im Gespräch erfolgen).
2. Außerdem wird es durch die Anreicherung mit Nutzerdaten möglich, die Daten automatisch miteinander zu verknüpfen. So lassen sich die Reichweite und die Vernetzung von 'Personal Learning Environments' (PLE, deutsch: Persönliche Lernumgebungen) und Portfolios erweitern.

Als grundlegende Struktur der Module des Lehr- und Lernbuches schlagen wir zunächst folgende aufeinander aufbauende Elemente vor:

1. Analyse großer Datenmengen in den Geisteswissenschaften – DH in unterschiedlichen Fachdisziplinen

Dieses Modul, das als Living Document angelegt wird, spezifiziert, was der Begriff Big Data Methoden in den Geisteswissenschaften allgemein und in einzelnen Fachdisziplinen bedeutet. Der aus den Wirtschafts- und Sozialwissenschaften entlehnte Terminus, der zunächst in seiner weitesten Bedeutung ("große Datenmengen betreffend") von den Geisteswissenschaften übernommen wurde, wird derzeit in zahlreichen Projekten genauer ausdifferenziert. Je nach Disziplin lässt sich eine große Bandbreite erkennen.

Während z.B. die Verbindung von Bildwissenschaften und Social Media Studies analytische Projekte hervorgebracht hat, die mehrere Millionen Datensätze einbeziehen, gelten in den Textwissenschaften meist schon Projekte, die mehrere Tausend Datensätze betrachten, als Vorhaben, die mit großen Datenmengen arbeiten. Unmittelbar an diese Beobachtung schließt sich die Frage an, was – übergreifend und disziplinspezifisch – als Daten oder Datensätze bezeichnet wird.

## 2. „Daten“ und große Datenmengen – (Ein-)Sichten aus den beteiligten Fächern

In diesem ebenfalls als Living Document konzipierten Abschnitt werden zunächst generische Definitionsansätze für Forschungsdaten aus DARIAH-DE aufgegriffen. Es bietet sich an, auf die Arbeit der DARIAH-DE AG 'Research Data Lifecycle' zurückzugreifen, da hier bereits eine sehr ausführliche Beschreibung von Forschungsdaten vorliegt. Die Lehrmittelsammlung bereichert die Arbeit der AG, indem fachspezifische Ansätze ergänzt werden, die durch Community Sourcing innerhalb und außerhalb von DARIAH-DE zusammengetragen werden können.

## 3. Datenmodelle und Modellierung von Daten

Die Arbeit mit großen Datenmengen erfordert die Entwicklung neuer Forschungsroutinen und Herangehensweisen. Die Modellierung von Einzelprojekten berücksichtigt darum direkt die Anforderungen des Forschungsgegenstandes. Dieser muss im Falle der Analyse großer Datenmengen zwar inhaltlich nicht anders beschaffen sein als Gegenstände traditioneller Forschung, grundsätzlich aber anders betrachtet werden. So unterscheiden sich die Fragestellungen erheblich von denen klassischer geisteswissenschaftlicher Projekte. Ein zentraler Punkt hierbei ist die Operationalisierung (vgl. Moretti 2013a) von Parametern und Konzepten. Die Modellierung von Forschungsprojekten trägt also dazu bei, einerseits Konzepte und Theorien geisteswissenschaftlicher Fächer neu zu überdenken und andererseits neue Fragestellungen und damit Herangehensweisen an Forschungsgegenstände zu entwickeln. Maßgeblich werden hier zunächst die Prozesse und Ergebnisse aus den Anwendungsbeispielen (Use Cases) zu dokumentieren sein. Daneben geht es aber auch um die Modellierung der Daten selbst, die so aufbereitet werden müssen, dass sie den Untersuchungsabsichten entsprechen, die sich je nach Fachdisziplin ergeben. Am Ende muss ein passendes Datenmodell gefunden werden, mit dem Informationen so aufbereitet werden können, dass sie zur Lösung der spezifischen Forschungsfrage beitragen.

## 4. Formalisierung

Vor allem bei der Verarbeitung großer Datenmengen können Formate der Daten und Metadaten sowie deren Einheitlichkeit eine große Bedeutung erlangen. Darum wird in diesem Modul auf die Forschungsarbeit von DARIAH-DE verwiesen, die sich mit der Formalisierung von Daten auseinandersetzt. Dazu gehört der Link zu den derzeit in Entwicklung befindlichen Empfehlungen zur Anwendung von Standards ebenso wie der Hinweis auf Lizenzen und rechtliche Rahmenbedingungen (vgl. Klimpel u.a. 2015), die für die (Wieder-)Veröffentlichung von Forschungsdaten wichtig sind.

## 5. Von der Forschungsfrage zum Ergebnis

Der Aufbau der Lehrmittelsammlung ist zweigeteilt: Zunächst wird der Gegenstand eines Einzelabschnitts, z.B. bestimmte Begriffe, Methoden oder Werkzeuge, vorgestellt, bevor schließlich einzelne Fachbereiche einbezogen werden. Auf diese Weise zusammengeführt, eröffnen die Inhalte der Lehrmittelsammlung eine disziplinspezifische Perspektive und ermöglichen einen interdisziplinären Vergleich oder eine transdisziplinäre Herangehensweise, die das Forschungsfeld der Digital Humanities stärker betont als die darin beheimateten Einzeldisziplinen. So ist es möglich, sowohl die oben erwähnten nach DH Kompetenz gestaffelten Nutzergruppen als auch die Ausprägungen der DH als Lehrdisziplin<sup>9</sup> zu berücksichtigen bzw. diese direkt anzusprechen. Dass Begriffe sowie Konzepte im Hinblick auf diverse Fachdisziplinen konkretisiert werden, ist ein Alleinstellungsmerkmal der Lehrmittelsammlung gegenüber anderen Lehrmitteln der Digital Humanities.

Ein hybrides Lehrbuch, das dieser Struktur entspricht,<sup>10</sup> ermöglicht sowohl den jeweiligen Fachcommunitys als auch den einzelnen NutzerInnen miteinander zu kooperieren und zu kollaborieren. Beiträge können eingereicht und dann kommentiert werden. Jedes "Kapitel" sieht einen Bereich zur fachdisziplinären Spezifikation vor, der kollaborativ von der jeweiligen Community gefüllt werden kann. Damit ist es möglich, die Fachwissenschaftler durch die Einbindung in den Helpdesk zu beteiligen.

---

9 Mit den einzelnen Ausprägungen der Digitalen Geisteswissenschaften als Lehrdisziplin beschäftigt sich etwa die "Arbeitsgruppe Referenzcurriculum Digital Humanities" (<http://www.dig-hum.de/arbeitsgruppe-referenzcurriculum-digital-humanities>). Zur Entwicklung der Geisteswissenschaften als Digital Humanities siehe auch Lauer 2013.

10 Als ein Beispiel für die Umsetzung könnte das Format des *Living Handbook* dienen. Die Struktur des *Living Handbook of Narratology*, das an der Universität Hamburg herausgegeben wird, entspricht einem ähnlichen, aber vereinfachtem Konzept. Siehe Hühn 2011.

## 5. Mediale Umsetzung der Lehrmaterialsammlung

Die mediale Umsetzung der Lehrmaterialsammlung muss den oben bereits aufgeführten Nutzerinteressen gerecht werden. Eine digitale Umsetzung der Lehr- und Lernunterlagen erscheint als zukunftssträchtigste Variante. Um die weiteren Entwicklungen innerhalb der Anwendungsbeispiele sowie die vorhandenen und noch zu entwickelnden Lehrmaterialien zu berücksichtigen, sollen die Lehrmittelsammlung sowie das Lehr- und Lernbuch multimedial umgesetzt werden. Das entstehende Lehrbuch soll sowohl für E-Reader, Tablets und Smartphones kompatibel sein. Außerdem wäre die Entwicklung einer App zu überlegen. Auf diese Weise ließe sich etwa besonders die wachsende Zielgruppe der „Digital Natives“ bedienen, die diese Form des Zugriffs auf Materialien am häufigsten nutzt (vgl. Ebner u.a. 2011).

Sowohl hinsichtlich der Umsetzung der Lehr- und Lernmaterialsammlung als auch des Lehr- und Lernbuchs muss gewährleistet sein, dass Beiträge aggregiert, Zusammenhänge hergestellt sowie in den eigenen Kontext eingeordnet werden können und die Beteiligung der Community möglich ist. NutzerInnen können sich allerdings nicht allein von den Empfehlungen zu ähnlich gelagerten und in einem Zusammenhang zu erschließenden Materialien leiten lassen, sondern müssen flexibel mit den neu entstehenden Szenarien umgehen.

Im Zuge der Umsetzung der Lehr- und Lernmittelsammlung wird nicht angestrebt, ein eigenes Lehrmanagementsystem (LMS) aufzubauen, das nach Bäumer u.a. (2004) aus den drei wesentlichen Säulen Administration, Kommunikation und Inhalt bestehen müsste. Obwohl der inhaltliche wie der kommunikative Aspekt durchaus im Sinne der oben dargelegten Strategie wäre, ist die administrative Funktionalität von LMS aus unserer Sicht noch zu sehr auf die Person des Lehrenden ausgerichtet, während unser Fokus eher auf den NutzerInnen als Lernenden liegt. Die Lehr- und Lernmittelsammlung wird – in Kombination mit dem Angebot des Lehr- und Lernbuches – als hybrides Format zwischen E-Portfolio-System und Persönlicher Lernumgebung (PLE) anzusiedeln sein. Der Aspekt der Administration, der dazu führen würde, dass ein LMS entsteht, wird bewusst zurückgestellt. Das Ziel eines solchen Hybrids ist es nicht, den NutzerInnen strenge Vorgaben zu machen, sondern sie darin zu unterstützen, ihren Lernprozess mithilfe der elektronischen Medien selbst zu organisieren, zu dokumentieren, zu reflektieren und zu präsentieren. Ein Vorteil ist, dass die elektronischen Profile der E-Portfolios leicht im Nachhinein bearbeitet und ergänzt werden können (vgl. Ebner/Maurer 2008). Damit sind sie auch für andere nützlich. Durch den Hybrid wird somit eine persönliche Lernumgebung geschaffen, die gleichzeitig prototypisch-generischen Charakter hat.

Die zu entwickelnde Sammlung lässt sich außerdem zur Dokumentation des Lernprozesses, u.a. auch zum „Lebenslangen Lernen“, einsetzen und hilft, Prozesse der Materialnutzung effektiver und effizienter zu gestalten. Durch die Umsetzung in digitalen Medien ergeben sich Mehrwerte. Hyperlinks und Annotationen können etwa direkt eingebunden werden, um auf weiterführende Informationsquellen zu verweisen und die einzelnen medialen Versionen (Webanwendung, App, E-Book) miteinander zu verbinden. Durch das Verschlagworten (Taggen) von Inhalten des E-Portfolios lassen sich außerdem ähnliche Informationsquellen anderer Anbieter einbinden.

In diesem Hybrid von E-Portfolio und PLE müssten darüber hinaus aufgenommen werden:

1. Technologien und Tools des Semantic-Webs: Die Informationen werden im WWW in einer strukturierten Form aufbereitet, die es Computern ermöglicht, die Inhalte zu verstehen und zu verarbeiten. Im Semantic-Web sind die Inhalte und deren Bedeutung von Maschinen interpretierbar und analysierbar (vgl. Softic u.a. 2013). Verknüpft und analysiert man die strukturierten Daten, lassen sich aus den bestehenden Daten automatisch neue Informationen generieren und neue Erkenntnisse gewinnen. Dadurch ist es zum Beispiel möglich, intelligente Systeme zu bauen, die passende Lernressourcen automatisch finden und den Lernenden zur Verfügung stellen.
2. Empfehlungsalgorithmen: Angesichts der rasch wachsenden Informationen, Dienste und Ressourcen im Web wird es für Lernende immer schwieriger, die passenden Informationen zu einem bestimmten Thema zu finden. Empfehlungssysteme können die erwünschten Informationen und Ressourcen vorschlagen und werden darum in vielen Bereichen, unter anderem in Lehr- und Lernsystemen, als Lösung gesehen (vgl. Taraghi u.a. 2013).

## 6. Ziele und Herausforderungen

### 6.1 Zielformulierung

Das hier vorgestellte Konzept umfasst sowohl die Lehr- und Lernmittelsammlung selbst als auch den Aufbau eines interaktiven Lehr- und Lernbuches. Das Material aus den Anwendungsfällen (Use Cases) im Rahmen von DARIAH-DE fließt in die Lehr- und Lernmittelsammlung ein und dient selbst als Gegenstand für die Erforschung verschiedener Fragestellungen (z.B. zu unterschiedlichen Szenarien zu Datenstandards, zur Generierung und Verarbeitung von Metadaten und deren Standards sowie zur Lizenzierung des Materials), die an konkreten Anwendungsfällen exemplarisch betrachtet werden können.

Ziel der Erstellung der Lehr- und Lernmittelsammlung zum Thema „große Datenmengen in den Geisteswissenschaften“ ist es, durch die modulare Strukturierung von Inhalten, die Entwicklungskosten und -zeiten möglichst gering zu halten und personelle Ressourcen eher dazu zu nutzen, das bereitgestellte Material durch eine persönliche Beratung zu vermitteln. Das bestehende Angebot frei zugänglicher Lehrmaterialien aus dem Forschungsprojekt DARIAH-DE I wird somit um einen personalisierten Vermittlungsansatz ergänzt. Die Struktur der Sammlung, die auf Kompetenzen aufbaut und mit Angeboten zu Workflows versehen werden soll, muss technisch so implementiert werden, dass sie sowohl den Ansprüchen von NutzerInnen genügt, die selbstständig nach Materialien suchen, als auch die Arbeit des DARIAH-DE Helpdesks qualitativ unterstützt.

Die Datenbank- bzw. Sammlungsarchitektur muss also Aspekte eines darbietungsorientierten Workflows und Lehrkonzepts abbilden und so gestaltet sein, dass sich Netzwerke zwischen NutzerInnen und Materialien, NutzerInnen und anderen NutzerInnen und vor allem NutzerInnen und der von DARIAH-DE bereitgestellten sozialen Infrastruktur bilden können. Dies lässt sich durch die offen angelegte Struktur des Materials in modularen Komponenten mit Verbindung zu anderen Sammlungen und der „Doing Digital Humanities Bibliografie“ gewährleisten. NutzerInnen sollen sich auf diese Weise einen Überblick über aktuelle Entwicklungen verschaffen und die Verbindungen zwischen einzelnen Themenfeldern, Ideen und Konzepten erkennen können (vgl. Arnold u.a. 2013, 110f.).

### 6.2 Qualitätssicherung

Einen wichtigen Faktor stellt die Qualitätssicherung dar. Dabei sind zwei Punkte wesentlich: Erstens gilt zu überprüfen, ob die jeweiligen Bildungsinhalte für bestimmte Lernsituationen wichtig sind; zweitens ist die fachliche und didaktische Aufbereitung der Materialien bedeutend und muss hinsichtlich ihrer Qualität begutachtet werden (vgl. Zauchner & Baumgartner 2007). Zusätzlich zu einem aufwändigen Peer-Review-Verfahren können UserInnen in das Qualitätsmanagement (QM) eingebunden werden, indem sie beispielsweise Kommentare, Annotationen und Empfehlungen zu den verwendeten Materialien hinzufügen. Dadurch ließe sich nachvollziehen, in welchen Bereichen bzw. Lernsituationen die Bildungsressourcen als sinnvoll und hilfreich erachtet werden (vgl. Hylén 2006). Auch wenn es nach Clements &

Pawlowski (2012) keine objektive Messung von Qualität gibt, lassen sich zwei verschiedene Ebenen unterscheiden, die Bestandteile des QM sein sollten:

1. Allgemeingültige Qualitätsansätze bieten Qualitätsmanagement- oder Qualitätssicherungsstrategien an, die unabhängig vom Einsatzgebiet sind. Ein Beispiel dafür ist die International Standards Organisation (ISO).
2. Spezifische Qualitätsinstrumente zielen darauf ab, Qualität für bestimmte Zwecke, wie etwa Qualitätssicherung von Metadaten, zu bewerten. Diese Instrumente sind manchmal in weitere Qualitätsansätze eingebettet, wie etwa beim Peer-Review-Verfahren.

### **6.3 Datenbank- und Sammlungsarchitektur**

Für die Sammlungs- und Datenbankarchitektur sollte über ein zentrales Content-Metadaten-Repository als Schnittstelle nachgedacht werden, um die Metadaten der Inhalte zu sammeln und sie den UserInnen zur Verfügung zu stellen. Dadurch ließe sich der Austausch der Materialien besser regeln. Sämtliche Metadaten müssten dazu einheitlich verwendet werden. Um dies zu gewährleisten, werden die innerhalb von DARIAH-DE entwickelten Richtlinien angewendet. Zusätzlich müssen die Metadaten, die ausgetauscht werden, dieselben Protokolle, wie z.B. OAI-PMH oder das SQL-Protokoll, berücksichtigen. Das Content-Metadaten-Repository würde zusätzlich gewährleisten, dass der Content-Metadaten-Provider automatisch Metadaten abrufen kann, die passend für die eigenen Lernobjekte sind. Jedes Repository bietet damit Informationen an, die mit den erforderlichen OAI-PMH Metadaten-Feldern Ort, Kennung und dem Datum der letzten Änderung beschrieben sind. Diese Informationen sind in den Learning Object Metadata abgebildet. Darüber hinaus erfolgt die Abbildung der internen Repository-Metadaten-Struktur in den LOM-Standard direkt beim Repository. Eine mögliche Suchanwendung fände anhand der Metadaten, die im Metadatenpeicher gespeichert sind, passende Lernobjekte. Die hier vorgeschlagene Architektur würde es den NutzerInnen ermöglichen, Lernressourcen von verschiedenen Repositorien zu finden, obwohl sie nur mit einer Suchanwendung arbeiten.

## 7. Literatur- und Webseitenverzeichnis

- „Arbeitsgruppe Referenzcurriculum Digital Humanities“, <http://www.dighum.de/arbeitsgruppe-referenzcurriculum-digital-humanities>. Letzter Zugriff: 14.12.2015.
- Arnold, Patricia; Kilian, Lars; Thilloßen, Anne u.a. 2013. *Handbuch E-Learning, Lehren und Lernen mit digitalen Medien*. 3. Aufl. Bielefeld: Bertelsmann Verlag.
- Bäumer, Michael; Malys, Bogusla & Michael Wosko. 2004. „Lernplattformen für den universitären Einsatz.“ In: Klaus Fellbaum & Marc Göcks (Hgg.). *eLearning an der Hochschule*. Aachen: Shaker Verlag. 121-140.
- Barker, Adam & Jonathan Stuart Ward. 2013. „Undefined by Data: A Survey of Big Data Definitions.“ *arXiv e-print*. New York: Cornell University Library. Online: <http://arxiv.org/abs/1309.5821>. Letzter Zugriff: 14.12.2015.
- Berners-Lee, Tim. 2009. „Linked Data - Design Issues.“ *World Wide Web Consortium (W3C)*. Online: <http://www.w3.org/DesignIssues/LinkedData.html>. Letzter Zugriff: 14.12.2015.
- Borgman, Christine. 2010. *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge, London: MIT Press.
- Clements, Kati & Jan Pawlowski. 2012. „User-Oriented Quality for OER: Understanding Teachers’ Views on Re-Use, Quality, and Trust.“ *Journal of Computer Assisted Learning* 28/1: 4-14. Online: <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2729.2011.00450.x/full>. Letzter Zugriff: 14.12.2015.
- Daconta, Michael; Obrst, Leo & Kevin Smith. 2003. *The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management*. Indianapolis, IN: Wiley.
- „DARIAH-DE“, <https://de.dariah.eu/>. Letzter Zugriff: 14.12.2015.
- „DARIAH-DE Sammlung von Schulungsmaterialien“, <https://de.dariah.eu/schulungsmaterial-sammlung>. Letzter Zugriff: 14.12.2015.
- Davies, John; Studer, Rudi & Paul Warren. 2006. *Semantic Web Technologies: Trends and Research in Ontology-based Systems*. Chichester: Wiley.
- „DH Mastermind“, <https://www.mindmeister.com/de/137418763/dh-projects-public-version-1>. Letzter Zugriff: 14.12.2015.
- „Doing Digital Humanities: A DARIAH-DE Bibliography“, <https://de.dariah.eu/bibliographie>. Letzter Zugriff: 14.12.2015
- Fensel, Dieter & Frank van Harmelen (Hgg.). *Towards the Semantic Web: Ontology-Driven Knowledge Management*. Indianapolis, IN: Wiley, 2003.

- Ebner, Martin & Hermann Maurer. 2008. „Can Microblogs and Weblogs change traditional scientific writing?“ In: Curtis J. Bonk, Mimi Miyoung Lee & Tom Reynolds (Hgg.). *Proceedings of E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2008*. Chesapeake, VA: AACE. 768-776.
- Ebner, Martin; Nagler, Walther & Martin Schön. 2011. „The Facebook Generation Boon or Bane for E-Learning at Universities?“ In: Theo Bastiaens & Martin Ebner (Hgg.). *Proceedings of EdMedia: World Conference on Educational Media and Technology 2011*. Chesapeake, VA: AACE. 3549-3557.
- Hühn, Peter u.a. (Hgg.). 2011. *the living handbook of narratology*. Hamburg: Hamburg University. Online: <http://www.lhn.uni-hamburg.de/>. Letzter Zugriff: 14.12.2015.
- Hylén, Jan. 2006. „Open Educational Resources: Opportunities and Challenges.“ *OECD's Centre for Educational Research and Innovation*. Online: <http://www.oecd.org/edu/ceri/37351085.pdf>. Letzter Zugriff: 14.12.2015.
- Klimpel; Paul & John H. Weitzmann. 2015. „Forschen in der digitalen Welt. Juristische Handreichung für die Geisteswissenschaften.“ *DARIAH-DE Working Papers* Nr. 12. Göttingen: DARIAH-DE. URN: urn:nbn:de:gbv:7-dariah-2015-5-0. Online: <http://webdoc.sub.gwdg.de/pub/mon/dariah-de/dwp-2015-12.pdf>. Letzter Zugriff: 14.12.2015.
- Laney, Doug. 2001. „3D-Data Management: Controlling Data. Volume, Velocity and Variety.“ *Gartner Blog Network*. Online: <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>. Letzter Zugriff: 14.12.2015.
- Lauer, Gerhard. 2013. „Die Digitale Vermessung der Kultur: Geisteswissenschaften als Digital Humanities.“ In: Heinrich Geiselberger & Tobias Moorstedt (Hgg.). *Big Data. Das neue Versprechen der Allwissenheit*. Berlin: Suhrkamp. 99-116.
- Lehr, Christian. 2011. *Web 2.0 in der universitären Lehre: Ein Handlungsrahmen für die Gestaltung technologiegestützter Lernszenarien*. Berlin: Freie Universität Berlin. Online: [http://www.diss.fu-berlin.de/diss/receive/FUDISS\\_thesis\\_000000035056](http://www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000035056). Letzter Zugriff: 14.12.2015.
- „L3T: Lehrbuch für Lernen und Lehren mit Technologien“, <http://www.l3t.eu/homepage>.
- Moretti, Franco. 2013. *Distant Reading*. London: Verso.
- Moretti, Franco. 2013a. „'Operationalizing': or, the function of measurement in modern literary theory.“ *Pamphlets of the Stanford Literary Lab* 6. Online: <http://litlab.stanford.edu/LiteraryLabPamphlet6.pdf>. Letzter Zugriff: 14.12.2015.
- Sahle, Patrick. 2013. „DH Studieren! Auf dem Weg zu einem Kern- und Referenzcurriculum der Digital Humanities.“ *DARIAH-DE Working Papers* Nr. 1. Göttingen: DARIAH-DE. URN: urn:nbn:de:gbv:7-dariah-2013-1-5. Online: <http://webdoc.sub.gwdg.de/pub/mon/dariah-de/dwp-2013-1.pdf>. Letzter Zugriff: 14.12.2015.

Softic, Selver; Taraghi, Benham; Ebner, Martin u.a. 2013. „Monitoring Learning Activities in PLE Using Semantic Modelling of Learner Behaviour.“ In: Andreas Holzinger, Martina Ziefle, Martin Hitz u.a. (Hgg.). *Human Factors in Computing and Informatics*. Heidelberg: Springer. 74-90.

“TaDiRAH (Taxonomy of Digital Research Activities in the Humanities)”, <https://github.com/dhtaxonomy/TaDiRAH>. Letzter Zugriff: 14.12.2015.

Taraghi, Benham; Ebner, Martin; Holzinger, Andreas u.a. 2013. „Web Analytics of user path tracing and a novel algorithm for generating recommendations in Open Journal Systems.“ *Online Information Review* 37/5: 672-691.

Zauchner, Sabine & Patrick Baumgartner. 2007. „Herausforderung OER (Open Educational Resources).“ In: Marianne Merkt, Kerstin Mayrberger, Rolf Schulmeister u.a. (Hgg.). *Studieren neu erfinden – Hochschule neu denken*. Münster: Waxmann Verlag. 244-252.