



Konzept Dissemination und Lehrmittelsammlung Cluster 5

Big Data in den Geisteswissenschaften (R 5.4.1)

Version 12.04.2016

Cluster 5

Verantwortliche Partner IEG Mainz, DHd/UniHH

DARIAH-DE **Aufbau von Forschungsinfrastrukturen für die e-Humanities**

Dieses Forschungs- und Entwicklungsprojekt wird / wurde mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF), Förderkennzeichen 01UG1110A bis N, gefördert und vom Projektträger im Deutschen Zentrum für Luft- und Raumfahrt (PT-DLR) betreut.

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

Projekt: DARIAH-DE: Aufbau von Forschungsinfrastrukturen für die e-Humanities

BMBF Förderkennzeichen: 01UG1110A bis N

Laufzeit: März 2011 bis Februar 2016

Dokumentstatus: <Final>

Verfügbarkeit: <DARIAH-DE-public>

Autoren:

<Mareike Höckendorff, DHd Hamburg>

<Marcus Held, IEG Mainz>

<Claudia Falk, IEG Mainz>

<Stefan Pernes, Uni Würzburg>

Revisionsverlauf:

Datum	Autor	Kommentare
27.08.2014	Mareike Höckendorff	Erste Notizen zu einem Disseminationskonzept von Cluster 5
14.10.2014	Marcus Held	Ergänzungen und Konkretisierung der Inhalte
16.11.2014	Mareike Höckendorff	weitere Ergänzungen
19.12.2014	Marcus Held	weitere Ergänzungen und erste Draft-Version
29.01.2015	Marcus Held	Einarbeitung von Kommentaren aus dem Cluster 5
24.03.2015	Mareike Höckendorff	Korrekturen und Kürzungen
09.04.2015	Stefan Pernes	Einarbeitung von Kommentaren aus dem Cluster 5 und Bereitstellung für das Konsortium
14.07.2015	Claudia Falk	Letzte Korrekturen und Veröffentlichung als Working Paper
12.04.2016	Stefan Pernes	Finalisierung

Inhaltsverzeichnis:

[Zu den Perspektiven dieses Papiers](#)

[Vom DARIAH-Helpdesk zu projektspezifischer Beratung](#)

[Inhalte der Lehrmittelsammlung](#)

[Darstellungsformen der Lehrmaterialsammlung](#)

[Mediale Umsetzung der Lehrmaterialsammlung](#)

[Ziele und Herausforderungen](#)

[Zielformulierung](#)

[Arbeitsschritte zur Zielerreichung](#)

[Qualitätssicherung](#)

[Datenbank- und Sammlungsarchitektur](#)

[Literatur](#)

Zu den Perspektiven dieses Papiers

Dieser Report setzt sich mit der Beschreibung von Inhalten, Darstellungsformen und der medialen Umsetzung einer Lehrmittelsammlung ausgehend vom Thema “Big Data Methodik in den Geistes- und Kulturwissenschaften” auseinander. Darüber hinaus wird ein Disseminationskonzept entwickelt, das die Vermittlung von Inhalten, Themen und Instrumenten dieses transdisziplinären Bereichs in den jeweiligen Fachdisziplinen beinhaltet. Das hier vorgestellte Konzept einer Lehr- und Lernmittelsammlung hat prototypisch-generischen Charakter und ist zunächst auf den Bereich der Nutzung und Anwendung von Big Data-Technologien und Methoden für geistes- und kulturwissenschaftliche Forschungsfragen ausgerichtet.

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 und Close Reading und eine Hinwendung zu Korpusanalysen und Distant Reading.¹ In den Bildwissenschaften unterstützen diese Methoden und Tools z.B. die Visualisierung und Untersuchung mehrerer oder gar mehrerer Tausend Bilder gleichzeitig. In den Geschichtswissenschaften können Datenbanken wie Wikimedia nach bestimmten biografischen Daten durchsucht und soziale Trends wie z.B. Entwicklung von Berufsbildern oder auch Lebenszeiterwartungen quantitativ ausgewertet werden. Für diese Methodik wird zwar oft auf den Begriff “Big Data” zurückgegriffen, tatsächlich werden aber in den Geistes- und Kulturwissenschaften Datenmengen verarbeitet, die nicht vergleichbar mit denen sind, die ursprünglich damit gemeint waren. Darum werden im Folgenden die Begriffe “Big Data Methodik” oder aber

¹ Moretti 2013.

“große Datenmengen” verwendet, um einerseits die Anlehnung an die impulsgebenden Trends beizubehalten, auf der anderen Seite aber auch eine gewisse Nähe zur tatsächlichen Datenbasis von Forschungsprojekten in den Geistes- und Kulturwissenschaften zu wahren.

Das vorliegende Disseminationskonzept berücksichtigt einerseits die sich durch die Anwendung von Big Data Methodik und den Einsatz datenbasierter Tools wandelnden Inhalte und Fragestellungen von Forschungsprozessen. Auf der anderen Seite wird bedacht, dass die Beschaffenheit der Daten in ihrer Vielfalt und Komplexität variieren kann und bestimmten fachbereichsspezifischen Entwicklungen unterliegen kann. Angesichts der erhöhten Anforderungen an Didaktik, Vermittlungsmethoden und -techniken, sowie der Variabilität und Komplexität der Daten, werden zwei prototypisch-generische Strategien entwickelt:

1. Erprobung quantitativer Analysemethoden anhand unterschiedlicher disziplinärer Perspektiven und Anforderungen wie sie in den Use Cases in Cluster 5 exemplarisch geleistet wird und
2. Analyse und Ergänzung der bisherigen Instrumente, Methoden und Dienste der Vermittlung von Methodenkompetenz durch DARIAH-DE.

Das Konzept der Lehr- und Lernmittelsammlung soll aus einer nutzerzentrierten² Perspektive heraus entwickelt werden und stellt eine Grundlage für die Erstellung einer Lehrmittelsammlung dar. Es wird empfohlen, im Anschluss an die Konzeptionierung der Disseminationsstrategie den Bedarf an Lehr- und Lernmitteln zur Big Data Methodik in Zusammenarbeit mit dem Stakeholdergremium Fachgesellschaften abzustimmen, sodass unterschiedliche disziplinäre Anforderungen berücksichtigt werden können.

Der hier entwickelte Ansatz entspringt dem Versuch, möglichst viele Fachbereiche und Anforderungsniveaus zumindest implizit zu berücksichtigen, um so die Akzeptanz in den jeweiligen geistes- und kulturwissenschaftlichen Fachcommunities zu erhöhen und die Verbreitung durch prototypische Angebote in Form von vorgelegten Fallstudien und angebotenen Veranstaltungen zu unterstützen. Durch die Angebote sollen sowohl der Einstieg in die Nutzung digitaler Tools und Methoden erleichtert werden, als auch neue Impulse der Weiterbildung für diejenigen gesetzt werden, die bereits mit Methoden der Digital Humanities arbeiten. Außerdem möchte das AP 5.4 auf methodologisch-reflexiver Ebene neue Impulse geben, indem neben der Erarbeitung und Bereitstellung des Materials den Nutzern ein Forum für den Austausch von Ideen angeboten wird. Dies kann einerseits durch die Organisation von Veranstaltungen und andererseits durch die Beteiligung einzelner Forscher an der Erstellung des Lehrbuches (s.u.) umgesetzt werden.

Das als intern verlinkte Sammlung von Tutorials zu Daten, Methoden und Tools konzipierte Lehrbuch basiert auf den Lehrmitteln im DARIAH-DE Portal, die die Big Data Methodik fokussieren und referenzieren. Dabei kommt die Taxonomy of Digital Research Activities in the Humanities (TaDiRAH)³ zur Anwendung und wird als Vorbild zur Taxonomisierung der anzulegenden Lehr- und Lernmittelsammlung herangezogen. Durch die Verwendung von

² Wobei mit Nutzer hier vornehmlich Geistes- und Kulturwissenschaftler gemeint sind, die in ihren Projekten bereits mit großen Datensammlungen arbeiten oder zukünftig arbeiten möchten.

³ <http://tadirah.dariah.eu/vocab/>

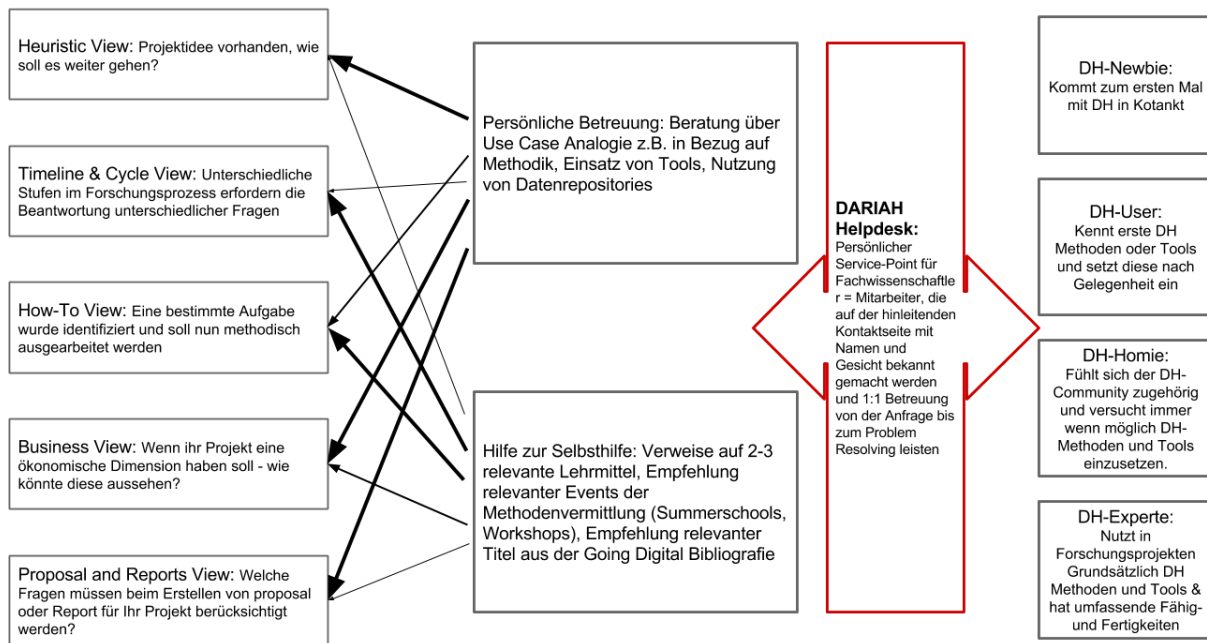
TaDiRAH wird es außerdem möglich, die Auszeichnung von Tool-Verzeichnissen über die Anreicherung bibliografischer Angaben mit der gleichzeitigen Reflexion der digitalen Forschungspraxis im Allgemeinen und im Hinblick auf das projektspezifische Anforderungsprofil im Besonderen zu verbinden, indem Annotationen zugelassen werden. So würde es möglich, bereits existierende, aber oft isoliert vorliegende Lehr- und Lernmittel sowohl generisch als auch projektspezifisch zu durchsuchen, zu vergleichen und zu nutzen. Auf diese Weise erlangt das Lehr- und Lernbuch nicht nur einen allgemein gehaltenen Charakter, sondern deckt auch spezifische Projektanwendungsfälle ab.⁴

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

Vom DARIAH-Helpdesk zu projektspezifischer Beratung

Die Idee eines DARIAH Helpdesk wird in einer eigens dafür konstituierten Service Unit (DeISU) ausgearbeitet und ist in Ansätzen bereits im Report 3.3.1 vom 24. Juni 2014 dargelegt worden. Die Hauptaufgabe der DeISU bestünde demnach in der Vermittlung persönlicher Berater, die nach Zuständigkeiten und Clusterzugehörigkeit erfolgen wird. Die Funktionalitäten eines solchen Helpdesk können wie folgt visualisiert werden:

⁴ Ein ähnliches Konzept der medialen Doppelläufigkeit verfolgte auch das Projekt von L3T: Lehrbuch für Lernen und Lehren mit Technologien(l3t.eu/homepage).



Innerhalb der Cluster erfolgt dann die persönliche Beratung über Use Case Analogien, die je nach Nutzer auf zwei Ebenen der Ansprache stattfinden können:

1. Weitere persönliche Beratung
2. Hilfe zur Selbsthilfe durch Verweise auf 2-3 relevante Lehr- und Lernmittel mittels der TaDiRAH-Taxonomie.

Das Beratungsrepertoire muss von einem Low-Level-Entry ausgehen und ist dann nach Nutzergruppen steigerbar; dabei stützt sich die Task Force auf die [DH Mastermind](#).

Wie Patrick Sahle in seinem Report "DH studieren"⁵ deutlich macht, ist von unterschiedlichen Ausbildungsprofilen auszugehen, die bei dem Aufbau einer Lehr- und Lernmittelsammlung besondere Berücksichtigung finden müssen. Zugleich muss ein niedrigschwelliger Einstieg gewährleistet werden, um auch breitere Zielgruppen, die bisher noch nicht mit Verfahren, Methoden und Instrumenten der DH vertraut sind, an die Materie heranzuführen. Zur Realisierung dieses Spagats soll die Lehr- und Lernmittelsammlung möglichst Materialien so aufbereiten, dass unterschiedliche Grade der Nutzung von DH-Methoden und Instrumenten Berücksichtigung finden. Außerdem sollen Materialien so gestaltet bzw. aufbereitet werden, dass verschiedene Zielgruppen angesprochen werden. Dazu zählen FachwissenschaftlerInnen, die DH-Verfahren entwickeln und/oder anwenden, fachwissenschaftliche und "Digitale GeisteswissenschaftlerInnen" / "Digital Humanists", Digitale FachwissenschaftlerInnen, Digitale GeisteswissenschaftlerInnen / Digital Humanists, und auch FachwissenschaftlerInnen ohne Erfahrungen in DH.⁶ Auch Absolventen von

⁵ [Sahle 2013](#).

⁶ Ebd., 6. Die Unterscheidung orientiert sich an der von Patrick Sahle entwickelten Heuristik von DH-Studierenden, die allerdings auch das Gesamtfeld der bisherigen Akteursszenarien treffend nachzeichnet.

DH-Programmen, die Sahle in sechs verschiedene Studiengruppen unterteilt,⁷ werden voraussichtlich verstärkt auf die Inhalte zugreifen und sicherlich mit sehr unterschiedlichen Erwartungen und Bedürfnissen eine Lehr- und Lernmittelsammlung im Bereich von Big Data Methoden nutzen.

Um die Komponenten zu nutzen und zu einem fallspezifischen bzw. individualisierten virtuellen "Work-/Textbook" zusammenzustellen, ließe sich auf das DARIAH-Repository zurückgreifen. So könnten die "Workbooks" als generische Lösungsstrategien abgelegt und für die Weiternutzung/-entwicklung zur Verfügung gestellt werden. Mit diesen Workbooks wären somit prototypisch-generische Lösungsstrategien für konkrete Anwendungsfälle dokumentiert. Diese könnten die Einstiegshemmungen für DH-Newbies senken und zugleich als Beitrag zur Projektberatung in das Helpdesk-System integriert werden. Dabei wäre auch zu überlegen, ob hierzu das DARIAH-Repository (Cluster 4) in seiner Grundstruktur genutzt werden kann.

In Cluster 5 werden derzeit drei Use Cases erprobt, die allesamt in textbasierten Fachdisziplinen verortet sind. Die schematische Darstellung zeigt, dass diese Use Cases 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 in "DH studieren!" definiert,⁸ werden von den Use Cases (jeweils in unterschiedlicher Intensität) berücksichtigt:

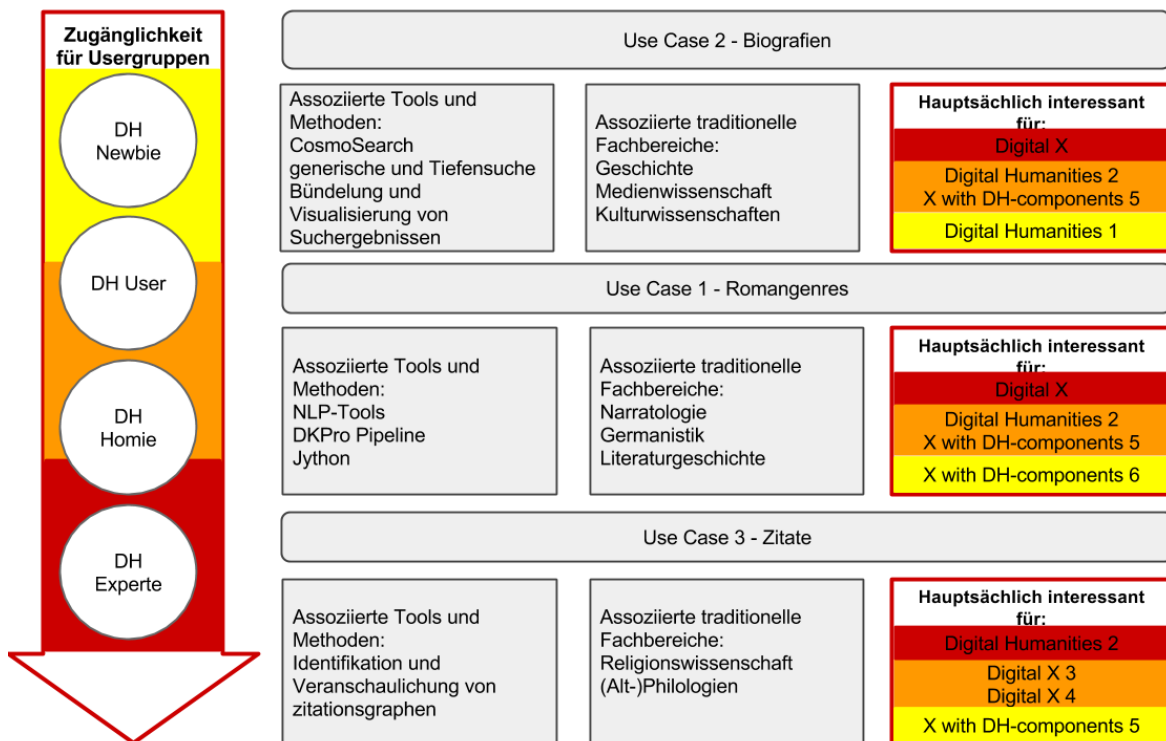


Abb. 2 Use Cases in Bezug auf Nutzergruppen und DH Studienrichtungen

Der Versuch, diese Use Case-zentrierte Darstellung mit der Idee eines DARIAH-Helpdesk zusammenzuführen, führt allerdings schnell zu der Einsicht, dass eine Entscheidung

⁷ Sahle 2013, S. 10f.

⁸ Sahle 2013, S. 10f.

bezüglich der Ausrichtung auf Nutzerinteressen gefunden werden muss. Die derzeitige Arbeit an den Use Case-spezifischen Lehr- und Lernmitteln 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 Nutzern erreicht. Abb. 1, die auch das Selbstverständnis der im Report 3.3.1 beschriebenen Service Unit aufgreift, zeigt jedoch ein anderes Verständnis von Nutzerorientierung. Hier tritt ein Nutzer an das Helpdesk heran, der mindestens zwei sich überlappende Interessen hat – ein fachliches und eines, das die Ausprägung der eigenen Methodenkompetenz betrifft. Handelt es sich z.B. um einen Literaturwissenschaftler, der noch neu innerhalb der DH ist und gerne ‘Natural Language Processing’ Tools auf ein Corpus aus Erzähltexten anwenden möchte, so müsste er entweder eine ausführliche persönliche Anleitung bekommen oder ein extra auf seinen Wissenstand abgestimmtes Lernmittel, da der Use Case an sich eher für fortgeschrittene Nutzer 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 also zuerst durch die DARIAH Service Unit (DeiSU) ein enger persönlicher Kontakt zu interessierten Forschern hergestellt. Im Zuge der persönlichen Beratung sollte zunächst erfragt werden, ob das Interesse sich eher auf die eigene Forschung oder auf die Lehre bezieht. Ist ersteres der Fall, so muss als nächstes erfragt werden, welches Vorwissen der Anfragende hat. Im Falle eines Forschers, der sich für Tools und Methoden des Natural Language Processing (Use Case 1) interessiert, beträfe dies ein Grundverständnis von Statistik und mathematischen Prozessen sowie von maschinellem Lernen. Hinzu kämen eventuell Vorkenntnisse in einer Programmiersprache. Je nach Kenntnisstand des Anfragenden kann auf Lernmittel der eigenen Sammlung verwiesen werden oder aber diese um Lektüreaktionen aus den entsprechenden Voraussetzungsgebieten ergänzt werden.

Ist ein Forscher nicht daran interessiert, die eigene Forschungsperspektive zu erweitern, sondern sucht nach Anregungen zur Vermittlung von Tools und Methoden der Digital Humanities, so benötigt er eine grundlegend andere Art der Beratung. Zwar ist auch hier zunächst zu erfragen, welche Grundkenntnisse im angefragten Gebiet vorhanden sind, allerdings ist eher davon auszugehen, dass nur Forscher DH lehren möchten, die zumindest ein Grundverständnis der angewandten Verfahren haben. Auch in diesem Fall sollte allerdings das Gebiet erfragt werden, über das gelehrt werden soll. Als Lösungsoption bietet sich in einem solchen Fall eher selten eines der DARIAH Lernmittel an. Stattdessen könnten dem Anfragenden bestimmte DARIAH Veranstaltungen empfohlen werden, die der Methodenkompetenzvermittlung dienen. Ist das Interesse breit gefächert, böte sich beispielsweise der Besuch der EDUROM Summerschool an, möchte sich der Anfragende auf ein spezielles Gebiet konzentrieren, so wird ihm der Besuch eines DARIAH Methodenworkshops weiterhelfen.

Inhalte der Lernmittelsammlung

Ausgehend von einer Beschreibung der in Cluster 5 bearbeiteten Use-Cases werden nun Unterschiede und Gemeinsamkeiten der Szenarien im Hinblick auf Forschungsprozesse, -methoden und -objekte herausgearbeitet. Im ersten Schritt werden die Szenarien als prototypische DH-Fragestellungen mit 'Big Data'-Methodik dargestellt, während es im zweiten Schritt gilt, generische Inhalte zu identifizieren, um sie als inhaltliche Struktur der Lehr- und Lernmittelsammlung zugrunde legen zu können.

- Der **Use Case 1** *Narrative Techniken und Untergattungen im deutschen Roman* demonstriert 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. Dabei kommen Verfahren zur automatischen Erkennung bestimmter Merkmale, wie zum Beispiel der Erkennung von Eigennamen oder von Passagen direkter Rede zum Einsatz. Die Merkmale werden im Anschluss als *Features* zueinander in Bezug gesetzt, um die Texte zu gruppieren und narratologische Kategorien nachprüfen zu können.

Ein dediziertes Deliverable des Use Cases ist es, die Entwicklung der Werkzeuge auch in Form von Lernmaterialien abzubilden, die verschiedene Einstiegspunkte bieten – auch für Anwender mit weniger Erfahrung im Umgang mit Programmiersprachen. Die technische Umsetzung teilt sich dabei in zwei Bereiche und entlastet die NutzerInnen: Sämtliche *NLP* Vorverarbeitungsschritte werden mit Hilfe des Darmstädter *DKPro Frameworks* als eine generische *Pipeline*, d.h. Abfolge aufeinander aufbauender Verarbeitungsschritte, umgesetzt und in Form eines effizienten Plaintextformats ausgegeben, das programmatisch einfach eingelesen und auch mit herkömmlicher Office Software geöffnet werden kann. Ab diesem Punkt beginnt die Weiterverarbeitung der Ergebnisse, im Use Case prototypisch umgesetzt anhand der Skriptsprachen *Python* und *R*.

Folgende Anforderungen ergeben sich dabei für die Benutzer: Die Werkzeuge des Use Case 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 von **Use Case 2** werden Verbindungen von individuellen historischen Lebensläufen und Internationalitätskriterien auf Grundlage von Wikipedia und mehreren europäischen Nationalbiografien untersucht und dabei 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 Quellen in mehreren Sprachen. Die Analyse biographischer Daten mittels semi-automatischer Erkennung von Korrelationen zwischen Personen, Orten und Ereignissen, ermöglicht schließlich die Erstellung personenbezogener Bewegungsprofile und die Aggregation der generierten Profile zur Unterstützung

spezifischer Fragestellungen. Zudem kann das System als wiederverwendbares Paket auch in anderen Kontexten eingesetzt werden, was allerdings eine spezifische Anpassung auf andere Datenbestände voraussetzt.

Folgende Anforderungen ergeben sich dabei für die Benutzer: 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 unter Umständen auch die Vermittlung einiger Schlüsselworte für komplexere Abfragen (eine *query language*). Ist es jedoch das Ziel der AnwenderIn, das System auf andere Datensätze oder inhaltliche Domänen anzupassen, so ergeben sich Anforderungen, die ähnlich gelagert sind wie in Use Case 1, diese unter Umständen jedoch noch übersteigen.

- Gegenstand von **Use Case 3** ist die Erschließung vollständiger diachroner Zitationsspuren in einer großen Sammlung griechischer und lateinischer Texte, die einen Zeitraum von der frühesten Kirchengeschichte bis in das 15. Jahrhundert abdecken. 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 bereit gestellt.

Folgende Anforderungen ergeben sich dabei für die Benutzer: Ähnlich wie für Use Case 2 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 skaliert. 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.

Um einen Überblick über bereits vorhandene Schulungsmaterialien, ihren inhaltlichen Aufbau und mögliche Pfadabhängigkeiten zu erhalten, wurde für den vorliegenden Report eine explorative Inhaltsanalyse einschlägiger Titel durchgeführt. Grundlage bilden sowohl die DARIAH-DE Sammlung von Schulungsmaterialien,⁹ als auch die Zotero-basierte Bibliografie *Doing Digital Humanities*.¹⁰ Es wurde dabei wie folgt vorgegangen: Alle relevanten Einträge – das heißt jene, die sich thematisch mit *Data Science* beschäftigen und als Schulungsmaterial geeignet sind – wurden gesichtet. Anschließend wurden die Einträge weiter nach Thema und

⁹ <https://de.dariah.eu/schulungsmaterial-sammlung>.

¹⁰ <https://de.dariah.eu/bibliographie>.

Schwierigkeitsgrad gefiltert und dabei bekannte und bewährte 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.

Anforderungsniveau A¹¹: *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.*

Anforderungsniveau B¹²:

- 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*

Anforderungsniveau C¹³:

- *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*
- *NLP*: *document indexing, document scoring, www scraping, building networks from text*

Anhand der Aufstellung werden auch Pfadabhängigkeiten ersichtlich, so zum Beispiel, dass ein aufgeklärter Einsatz von *Machine Learning*-Verfahren zumindest Grundlagen aus Statistik und Programmierung benötigt. Ein Blick in die betreffenden Lehr- und Lernmittel führt zu folgenden Hypothesen hinsichtlich ihres internen inhaltlichen Aufbaus: Sowohl in der ersten, als auch in der dritten Gruppe 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 beginnende, als auch für fortgeschrittene AnwenderInnen, die eine solche

¹¹ Titel zu A: Universität Würzburg - DigiVoy: TextGrids Digitale Bibliothek mit Voyant entdecken, Sinclair et al. - Voyant Screencasts, Graham et al. - The Historian's Macroscopic, Schulserver Hessen - SQL-Tutorial, The Programming Historian - Understanding Regular Expressions.

¹² Titel zu B: Khan Academy - Probability and statistics, Haslwanter - Introduction to Statistics using Python, Groß - R Reader. Arbeiten mit dem Statistikprogramm R, The Programming Historian - Introduction to Python, Hochschule der Medien Stuttgart - Einführung in Python.

¹³ Titel zu C: Coursera - Machine Learning, Hochschule der Medien Stuttgart - Data Mining mit Scikit Learn, Riddell - Text Analysis with Topic Models for the Humanities and Social Sciences, Karsdorp - Python Programming for the Humanities, Layton - Text mining online data with scikit-learn.

Struktur bedingen. Demgegenüber ist in der zweiten Gruppe, deren Inhalte relativ klar abgegrenzte Grundlagen vermitteln, am ehesten ein lückenloser Gerüstbau vorzufinden.

Ausgehend von der hier untersuchten Stichprobe ist festzustellen, dass es - insbesondere für höhere Anforderungsniveaus - *Best Practice* ist, weiterführende Hinweise auf sämtliche für das Verständnis notwendigen Inhalte anzubieten und auch auf Nebenaspekte, wie die Einrichtung der Arbeitsumgebung oder alternativer Programmpakete, einzugehen. Somit liegt nahe, bei der Herstellung einer Use Case-Dokumentation einen "roten Faden" zu definieren, 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 Use Cases, gilt es, auch das bereits in DARIAH I erstellte und gesammelte Material in Bezug 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 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 bereitgestellt werden kann. Ergänzend soll die durch Cluster 5 bearbeitete erweiterte Bibliografie "Doing Digital Humanities" herangezogen und mit Inhalten und Materialien aus DARIAH I und DARIAH II verbunden werden, wodurch die Use Cases aus Cluster 5 in einen Gesamtzusammenhang gebracht werden. Die so entstehende Bibliografie zu "Big Data Methoden" stellt ihrerseits einen Beitrag zur Selbsthilfe dar, die für die Arbeit des Helpdesks der DeISU - wie oben skizziert - von zentraler Bedeutung sein wird.

Die bisherige Sammlung soll themenspezifisch in die aufzubauende Daten- und Sammlungsarchitektur integriert werden und stellt damit selbst einen Anwendungsfall in Bezug auf die Integration und Evaluierung von Lehr- und Lernmaterialien im Schritt zu OERs dar. Daneben stellen die vorhandenen Daten und Materialien der Use Cases auch in Bezug auf die Anreicherung mit Metadaten und Standards, Bewertung der Qualität, die Klärung von lizenzrechtlichen Fragen im Hinblick auf die 4 R's der OER (*re-use, re-vise, re-mix und re-distribute*) einen wichtigen Anwendungsfall dar. Kategorisierung und Taxonomisierung führen die vorhandenen Wissensressourcen zusammen und werden durch ein integratives Komponentensystem in der Datenbank- und Sammlungsarchitektur weiterentwickelt. Hierbei wird die vorhandene Taxonomie TaDiRAH¹⁴ eingesetzt und weiterentwickelt, 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 holistisch 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 mit Blick auf unterschiedliche fachwissenschaftliche Perspektiven. Bei der Konzeption der Lehr- und Lernmaterialsammlung sowie der einzelnen Lehr- und Lernmittel sollte also berücksichtigt

¹⁴ <https://github.com/dhtaxonomy/TaDiRAH>.

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.

Darstellungsformen der Lehrmaterialsammlung

Entsprechend des Selbstverständnisses von DARIAH-DE wird die Lehr- und Lernmittelsammlung als Infrastruktur des methodenorientierten Wissenstransfers und der Methodenkompetenzvermittlung verstanden. Innerhalb einer so konzipierten Sammlung kommen den Verknüpfungen einzelner Lehr- und Lernmittel miteinander eine besondere Rolle zu. Die Sammlung sollte so gestaltet werden, dass einerseits eine thematische Fortentwicklung von basalen zu spezifischen Kompetenzen ermöglicht wird, andererseits Nutzer aber auch Querverbindungen zwischen thematisch verbundenen Einzelabschnitten folgen bzw. diese selbst aufgrund von spezifischen Aktivitätsmustern und Workflows herstellen können. Eine derartige Konzeption wurde hier beispielhaft für die bereits teilweise vorhandenen oder gerade in Planung befindlichen Lehr- und Lernmittel des Cluster 5 in Ergänzung mit den Lehr- und Lernmaterialien zum Thema "Big Data Methoden in den Geisteswissenschaften", die bereits in der ersten Projektphase von DARIAH-DE erarbeitet wurden, dargelegt. Für Aufbau und Weiterentwicklung einer solchen Lehr- und Lernmittelsammlung wird vorgeschlagen, ein zweigliedriges Verfahren zu verfolgen:

1. Erstellung einzelner Lernmittel auf Grundlage der Use-Cases,
2. Zusammenstellung und Herausgabe eines intermedial verfügbaren Lehr- und Lernbuches, welches von DARIAH-Mitarbeitern und Vertretern der Fachgesellschaften, sowie ausgesuchten WissenschaftlerInnen erarbeitet wird und dann potentiell durch Nutzer ergänzt werden kann.

Beide Herangehensweisen folgen den Zielen und 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 Modularisierung, die durch Annotieren und Kommentieren einzelner Komponenten und Module zusätzlich unterstützt werden kann, ist eine Lösungsmöglichkeit, verschiedenen Nutzerinteressen entgegenzukommen und zugleich ein höchstmögliches Maß an Effizienz, Effektivität und Flexibilität zu ermöglichen. 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 des Individualnutzers anzupassen. Dabei ist von besonderem Interesse, inwiefern sich Materialien oder einzelne Bestandteile in verschiedenen

Projektzusammenhängen wiederverwenden lassen, da dies ein konkreter Indikator für die Skalierbarkeit¹⁵ 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 derzeit im Rahmen der AG "Research Data Lifecycle" ausgearbeitet werden.

Die Aufteilung eines Moduls in Komponenten ermöglicht es auf der einen Seite, die Aktualität einzelner Lerninhalte zu steigern, da aktualisierungsbedürftige Bestandteile einzeln angepasst werden können. Auf der anderen Seite 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 den Nutzer als Lehrend-Lernenden zu begreifen, der darüber hinaus daran mitwirkt, die Inhalte der Lehr- und Lernmaterialsammlung in ein Linked Open Data (LOP)¹⁶ System zu überführen. 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.¹⁷

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

- I. 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 Möglichkeiten mit anderen Nutzern zu kooperieren und zu kollaborieren (die Kommunikation kann dabei semi-automatisch über Kommentare und Annotationen oder direkt im Gespräch erfolgen).
- II. Außerdem wird es durch die Anreicherung mit Nutzerdaten möglich, die Daten automatisch miteinander zu verknüpfen. So lassen sich die Reichweite und die Vernetzung von PLEs und Portfolios erweitern.

¹⁵ Vgl. Lehr 2011, S. 61ff.

¹⁶ Vgl. Tim Berners-Lee, Linked Data - Design Issues, <http://www.w3.org/DesignIssues/LinkedData.html>

¹⁷ Vgl. John Davies, Rudi Studer, Paul Warren: *Semantic Web Technologies: Trends and Research in Ontology-based Systems* Wiley (2006); ders., Dieter Fensel, Frank van Harmelen: *Towards the Semantic Web: Ontology-Driven Knowledge Management*. Wiley (2003); Michael C. Daconta, Leo J. Obrst, Kevin T. Smith: *The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management*. Wiley (2003).

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 als Living Document konzipierte Modul 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 als "große Datenmengen betreffend" von den Geisteswissenschaften übernommen wurde, wird derzeit in zahlreichen Projekten genauer ausdifferenziert. Dabei zeigt sich eine Bandbreite, die vor allem disziplinspezifisch variabel ist. Während z.B. die Verbindung von Bildwissenschaften und Social Media Studies analytische Projekte hervor gebracht hat, die mehrere Millionen Datensätze einbeziehen, werden in den Textwissenschaften meist schon Projekte, die mehrere Tausend Datensätze betrachten als mit großen Datenmengen operierend bezeichnet. 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 AG 'Research Data Lifecycle' zurückzugreifen, da hier bereits eine sehr ausführliche Beschreibung dessen vorliegt, was Forschungsdaten im Allgemeinen ausmacht. Der dezidierte Mehrwert, den die Lehrmittelsammlung der Arbeit der AG hinzufügt, ist auch hier wieder die Ergänzung um fachspezifische Ansätze, die durch Community Sourcing innerhalb und außerhalb von DARIAH-DE zusammengetragen werden können.

3. Datenmodelle und Modellierung von Daten

Die Betrachtung großer Datenmengen erfordert die Entwicklung neuer Forschungsroutinen und Herangehensweisen. Die Modellierung von Einzelprojekten antwortet darum direkt auf die Anforderungen des Forschungsgegenstandes, der im Falle der Analyse großer Datenmengen zwar inhaltlich nicht anders sein muss als Gegenstände traditioneller Forschung, der aber grundsätzlich anders betrachtet werden muss. So unterscheiden sich auch die Fragestellungen nicht unerheblich von denen klassischer geisteswissenschaftlicher Projekte. Ein zentraler Punkt hierbei ist die Operationalisierung¹⁸ von zu betrachtenden Parametern und Konzepten. Die Modellierung von Forschungsprojekten trägt also nicht unerheblich dazu bei, dass einerseits Konzepte und Theorien geisteswissenschaftlicher Fächer neu überdacht

¹⁸ Vgl. Moretti "Operationalizing": or, the function of measurement in modern literary theory", <http://litlab.stanford.edu/LiteraryLabPamphlet6.pdf>, Stand 19.3.2015.

werden und andererseits neue Fragestellungen und damit Herangehensweisen an Forschungsgegenstände entwickelt werden. Maßgeblich werden hier die Prozesse und Ergebnisse zunächst aus den 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 oder geschaffen 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 und 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, die für die (Wieder-)Veröffentlichung von Forschungsdaten wichtig sind.

5. Von der Forschungsfrage zum Ergebnis

Der Aufbau der Lehrmittelsammlung folgt einem binären Ansatz: zunächst wird der Gegenstand eines Einzelabschnitts generisch zu fassen versucht – z.B. indem bestimmte Begriffe, Methoden oder Werkzeuge vorgestellt werden –, bevor schließlich einzelne Fachbereiche fokussiert werden. Auf diese Weise zusammengeführt, ermöglichen die Inhalte der Lehrmittelsammlung immer eine disziplinspezifische Perspektive, ein interdisziplinärer 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 Usergruppen als auch die von Lauer beschriebenen Ausprägungen der DH als Lehrdisziplin zu berücksichtigen bzw. diese direkt anzusprechen. Dabei kann vor allem die multiple Fachdisziplinen einbeziehende Konkretisierung der Begriffe und Konzepte als Alleinstellungsmerkmal der so entstehenden Lehrmittelsammlung gegenüber anderen Lehrmitteln der Digital Humanities betrachtet werden.

Ein hybrides Lehrbuch, das dieser Struktur entspräche,¹⁹ ermöglicht sowohl den jeweiligen Fachcommunitys als auch den einzelnen Nutzern miteinander zu kooperieren und zu kollaborieren. Beiträge können eingereicht und dann kommentiert werden. Jedes "Kapitel" sieht dabei einen Bereich zur fachdisziplinären Spezifikation vor, der kollaborativ von der jeweiligen Community (ausgehend vom Stakeholdergremium, das vom DHd in Hamburg koordiniert wird) gefüllt werden kann. Damit ist es möglich, die Fachwissenschaftler sowohl durch die Einbindung in das Helpdesk als auch über das Stakeholdergremium "Fachgesellschaften" als "Ideenstakeholder" zu beteiligen. Die Nutzung der digitalen

¹⁹ 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.: Hühn 2011.

Infrastruktur von DARIAH-DE erlaubt es, entweder frei im Internet oder auch restriktiv im Intranet, eine kollaborative Entwicklungsumgebung zu schaffen.

Mediale Umsetzung der Lehrmaterialsammlung

Die mediale Umsetzung der Lehrmaterialsammlung muss den oben bereits aufgeführten Nutzerinteressen gerecht werden. Eine hybride Umsetzung der digitalen Lehr- und Lernunterlagen erscheint als zukunftsfähigste Variante. Um die weiteren Entwicklungen der Use Cases sowie die Zusammenführung von schon vorhandenem und zu entwickelndem Lehrmaterial zu berücksichtigen, soll bei der Anlage der Lehrmittelsammlung (sowohl mit der Sammlung selbst, als auch mit dem Lehr- und Lernbuch) eine multiperspektivische Darstellungsform umgesetzt werden. Das entstehende Lehrbuch soll sowohl für E-Reader als auch für Tablets und Smartphones kompatibel sein bzw. könnte zusätzlich in Form einer App weiterentwickelt werden. So ließe sich vor allem die wachsende Zielgruppe der "Digital Natives" bedienen, die diese Form des Zugriffs auf Materialien am häufigsten nutzt.²⁰ 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. Dabei ist allerdings nicht auszuschließen, dass Nutzer sich von den Empfehlungen zu ähnlich gelagerten und in einem Zusammenhang zu erschließenden Materialien nicht allein leiten lassen können, sondern flexibel mit den neu entstehenden Szenarien umgehen müssen.

Im Zuge der Umsetzung der Lehr- und Lernmittelsammlung wird nicht angestrebt, ein eigenes Lehrmanagementsystem aufzubauen, das nach Bäumer et al. (2004) aus den drei wesentlichen Säulen Administration, Kommunikation und Inhalt bestehen müsste. Obwohl sowohl der inhaltliche als auch 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 Nutzern als Lernende liegt. Die Lehr- und Lernmittelsammlung wird in Kombination mit dem Angebot des Lehr- und Lernbuches in der medialen Umsetzung eines Hybrids zwischen E-Portfolio-System und Persönlicher Lernumgebung (PLE) anzusiedeln sein. Der Aspekt der Administration, der dazu führen würde, dass ein LMS entsteht, wird jedoch bewusst zurückgestellt. Das Ziel eines solchen Hybrids ist es nicht, den Nutzern 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.²¹ Damit sind sie auch für andere nützlich. Durch den Hybrid wird somit eine persönliche Lernumgebung geschaffen, die gleichzeitig prototypisch-generischen Charakter hat.

²⁰ Ebner, Martin; Nagler, Walther; Schön, Martin (2011). The Facebook Generation Boon or Bane for E-Learning at Universities?. In: World Conference on Educational Multimedia, Hypermedia and Telecommunications; 2011, S. 3549-3555

²¹ Vgl. Ebner & Maurer 2008.

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

Grundlegend müssten in diesem Hybrid von E-Portfolio und PLE darüber hinaus Berücksichtigung finden:

- a. Technologien und Tools des Semantic-Webs. Dabei werden die Informationen im WWW in einer strukturierten Form so aufbereitet, dass es Computern ermöglicht wird, sie inhaltlich zu verstehen und zu verarbeiten. Im Semantic-Web sind die Inhalte und deren Bedeutung also von Maschinen interpretierbar und analysierbar.²² Verknüpft und analysiert man die strukturierten Daten, lassen sich aus den bestehenden Daten automatisiert neue Informationen generieren und neue Erkenntnisse gewinnen. Dadurch ist es zum Beispiel möglich, intelligente Systeme zu bauen, die passende Lernressourcen automatisiert ohne menschliches Eingreifen finden und den Lernenden zur Verfügung stellen.
- b. Es sollten Empfehlungsalgorithmen integriert werden: 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. Es liegt nahe, dass Systeme die erwünschten Informationen und Ressourcen vorschlagen sollten. Empfehlungssysteme werden in vielen Bereichen, unter anderem in Lehr- und Lernsystemen, als Lösung dieser Probleme gesehen.²³

Ziele und Herausforderungen

Zielformulierung

Das hier vorgestellte Konzept umfasst sowohl die Lehr- und Lernmittelsammlung selbst, als auch den Aufbau eines interaktiven Lehr- und Lernbuches. Das aus den Use Cases des Clusters 5 generierte Material fließt in die Lehr- und Lernmaterialsammlung ein, dient aber zugleich selbst der Erforschung verschiedener Fragestellungen (z.B. unterschiedliche Szenarien zu Datenstandards, der Generierung und Verarbeitung von Metadaten und deren Standards, sowie der 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, Entwicklungskosten und -zeiten möglichst gering zu halten und personelle Ressourcen eher so zu nutzen, dass das bereitgestellte Material durch eine persönliche Beratung vermittelt

²² Vgl. Softic et al. 2013.

²³ Vgl. Taraghi et al. 2013.

wird. Das bestehende Angebot frei zugänglicher Lehrmaterialien aus DARIAH 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 Nutzern genügt, die selbstständig nach Materialien suchen, als auch die Arbeit der DeISU qualitativ unterstützt.

Die Datenbank- bzw. Sammlungsarchitektur muss also sowohl Aspekte eines darbietungsorientierten und instruktional konzipierten Workflows abbilde, als auch im Sinne des Konnektivismus so gestaltet sein, dass sich Netzwerke zwischen Nutzern und Materialien, Nutzern und anderen Nutzern und vor allem Nutzern und der von DARIAH-DE bereitgestellten sozialen Infrastruktur bilden können. Dies soll durch die offen angelegte Struktur des Materials in modularen Komponenten mit Verbindung zu anderen Sammlungen und der "Doing Digital Humanities Bibliografie" gewährleistet werden. Am Ende soll es Nutzern möglich sein, sich einen Überblick über aktuelle Entwicklungen zu verschaffen und die Verbindungen zwischen einzelnen Themenfeldern, Ideen und Konzepten zu erkennen.²⁴

Arbeitsschritte zur Zielerreichung

Folgende Arbeitsschwerpunkte sind im AP 5.4 bis zum Ende der Laufzeit von DARIAH II im Februar 2016 vorrangig zu berücksichtigen:

- Bearbeitung der Use Cases als prototypische Anwendungsfälle für Big Data Methoden in den Geisteswissenschaften,
- Erstellung von Lehrmaterialien für die in den Use Cases genutzten Tools und Methoden sowie Aufbereitung der Use Cases als Beispielszenarien für deren Verwendung,
- Durchführung von Workshops:
 - Expertenworkshops zu Use Cases und darin verwendeten Technologien und Methoden,
 - Methodenworkshops zur Vermittlung der in den Use Cases angewendeten Verfahren,
 - Stakeholderworkshop zur Thematik digitaler Methoden in den Geisteswissenschaften.

Qualitätssicherung

Einen bedeutenden, noch näher zu bestimmenden Faktor stellt die Qualitätssicherung dar. Dabei sind zwei Punkte wesentlich. Einerseits ist zu überprüfen, ob die jeweiligen Bildungsinhalte für bestimmte Lernsituationen wichtig sind. Zudem ist die fachliche und didaktische Aufbereitung der Materialien bedeutend und muss hinsichtlich ihrer Qualität begutachtet werden.²⁵ Zusätzlich zu einem aufwändigen Peer-Review-Verfahren können User/innen mit in das Qualitätsmanagement (QM) eingebunden werden. Sie könnten einen wesentlichen Beitrag zum QM leisten, indem sie beispielsweise Kommentare, Annotationen

²⁴ Vgl. Arnold et al. 2013, 110f.

²⁵ Vgl. Zauchner & Baumgartner 2007.

und Empfehlungen über die verwendeten Materialien hinzufügen. Dadurch ließe sich nachvollziehen, in welchen Bereichen bzw. Lernsituationen die Bildungsressourcen als sinnvoll und hilfreich erachtet wurden.²⁶ Auch wenn es nach Clements & Pawlowski (2012) keine objektive Messung von Qualität gibt, können zwei verschiedene Ebenen unterschieden werden, 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.

Datenbank- und Sammlungsarchitektur

Bei der Sammlungs- und Datenbankarchitektur müsste über ein zentrales Content-Metadaten-Repository als Schnittstelle nachgedacht werden, um einerseits die Metadaten der Inhalte zu sammeln, andererseits auch um diese für die User/innen zur Verfügung zu stellen. Dadurch könnte der Austausch der Materialien besser geregelt werden. Dazu ist es aber notwendig, dass sämtliche Metadaten von allen 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, die – einander ergänzend – eingesetzt werden könnten. 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 Metadaten Speicher gespeichert sind, passende Lernobjekte. Die hier vorgeschlagene Architektur würde es den Nutzer/innen ermöglichen, Lernressourcen von verschiedenen Repositories zu finden, obwohl sie nur mit einer Suchanwendung arbeiten.

Literatur

Arnold, P., Kilian, L., Thillosen, A., Zimmer, G.: Handbuch E-Learning, Lehren und Lernen mit digitalen Medien. 3. Aufl. Bielefeld: Bertelsmann Verlag, 2013.

²⁶ Vgl. Hylén 2006.

Bäumer, M., Malys, B., Wosko, M.: Lernplattformen für den universitären Einsatz. In: K. Fellbaum & M. Göcks (Hrsg.), eLearning an der Hochschule, Aachen: Shaker Verlag, 2004, 121-140.

Barker, Adam; Ward, Jonathan Stuart: Undefined by Data. A survey of Big Data definitions. ArXiv e-print, September 2013.

Berners-Lee, Tim: Linked Data - Design Issues.
<http://www.w3.org/DesignIssues/LinkedData.html>, Stand 31.3.2015 13:21 Uhr.

Borgmann, Christine: Scholarship in the Digital Age: Information, Infrastructure, and the Internet. Cambridge & London: MIT Press 2010.

Clements, K. I. & Pawlowski, J.M.: *User-Oriented Quality for OER: Understanding Teachers' Views on Re-Use, Quality, and Trust*. In: Journal of Computer Assisted Learning, 28/2012, 1, 4-14. <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2729.2011.00450.x/full>, Stand 1.4.2015, 9:19 Uhr.

Daconta, Michael C.; Obrst, Leo J.; Smith, Kevin T.: *The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management*. Wiley (2003).

Davies, John; Studer, Rudi; Warren, Paul: *Semantic Web Technologies: Trends and Research in Ontology-based Systems*. In: Fensel, Dieter; Harmelen, Frank van: *Towards the Semantic Web: Ontology-Driven Knowledge Management*. Wiley (2003)

Ebner, M. & Maurer, H. (2008). Can Microblogs and Weblogs change traditional scientific writing ? . In: Proceedings of E-Learn 2008, Las Vegas, , Chesapeake, VA: AACE , 768-776.

Ebner, M., Nagler, W., Schön, M. (2011). The Facebook Generation Boon or Bane for E-Learning at Universities?. In: World Conference on Educational Multimedia, Hypermedia and Telecommunicati-ons; 2011, S. 3549-355

Hühn, Peter et al. (eds.) (2011). *the living handbook of narratology*. Hamburg: Hamburg University. URL = <http://www.lhn.uni-hamburg.de/>, Stand 1.4.2015, 9:23 Uhr.

Hylén (2006). *Open Educational Resources: Opportunities and Challenges*. OECD's Centre for Educational Research and Innovation, 2006. <http://www.oecd.org/edu/ceri/37351085.pdf>, Stand 1.4.2015, 9:22 Uhr.

Laney, Doug: 3D-Data Management. Controlling Data. Volume, Velocity and Variety. <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-D ata-Volume-Velocity-and-Variety.pdf>, Stand 31.3.2015 13:16 Uhr.

Lehr, C. (2011): Web 2.0 in der Universitären Lehre. Ein Handlungsrahmen für die Gestaltung technologiegestützter Lernszenarien, Dissertationsschrift an der Fakultät für Wirtschaftswissenschaft der Freien Universität Berlin, http://www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000035056. Stand 1.4.2015, 9:04 Uhr

Moretti, Franco: "Operationalizing": or, the function of measurement in modern literary theory", <http://litlab.stanford.edu/LiteraryLabPamphlet6.pdf>, Stand 19.3.2015

Moretti, Franco: Distant Reading. London, Verso 2013.

Sahle, P. (2013). DH studieren!

<http://webdoc.sub.gwdg.de/pub/mon/dariah-de/dwp-2013-1.pdf> ..., Stand 31.3.2015 13:19 Uhr.

Softic, S., Tarahi, B., Ebner, M., De Vocht, L., Mannens, E., Van De Walle, R.: Monitoring Learning Activities in PLE Using Semantic Modelling of Learner Behaviour. In: A. Holzinger; M. Ziefle; M. Hitz & M. Debevc (Hrsg.). Human Factors in Computing and Informatics. Berlin/Heidelberg: Springer, 2013, 74-90.

Taraghi, B., Ebner, M., Holzinger, A., Grossegger, M.: Web Analytics of user path tracing and a novel algorithm for generating recommendations .In Open Journal Systems, 2013. Online Information Review. akzeptiert, im Druck.

Zauchner, S. & Baumgartner, P.: *Herausforderung OER (Open Educational Resources)*. In: M. Merkt, K. Mayrberger, R. Schulmeister, A. Sommer, I. van den Berk (Hrsg.) Studieren neu erfinden – Hochschuleneu denken (S. 244 – 252). Münster: Waxmann Verlag, 2007.