Forschungsinfrastrukturen in den digitalen Geisteswissenschaften.

Wie verändern digitale Infrastrukturen die Praxis der Geisteswissenschaften?

Martin Huber, Sybille Krämer, Claus Pias

Symposienreihe „Digitalität in den Geisteswissenschaften“

Gefördert durch

DFG Deutsche Forschungsgemeinschaft

IMPRESSUM

HERAUSGEBER
Martin Huber, Sybille Krämer, Claus Pias

KONTAKT
Julia Menzel
Digitalität in den Geisteswissenschaften
DFG-geförderte Symposienreihe
Universität Bayreuth
Universitätsstr. 30
95447 Bayreuth
www.digitalitaet.dfg@uni-bayreuth.de

1. Auflage November 2019

Wir danken der Deutschen Forschungsgemeinschaft (DFG Projekt Projektnummer 287972711) für die Förderung.

Inhalt

Martin Huber: Einleitung……3

I. Grundlagen

Heike Neuroth, Ulrike Wuttke: Eine gemeinsame geisteswissenschaftliche Forschungsdatenkultur: Utopie oder Dystopie?……8

Gerhard Lauer: Ist da eine Veränderung in den Geisteswissenschaften? Infrastrukturen und ihre Folgen für die Praxis der Geisteswissenschaften……27

II. Beispiele aus den Fächern

Hanno Ehrlicher, Jörg Lehmann: Datenerhebung als epistemologisches Labor – Überlegungen am Beispiel der virtuellen Forschungsumgebung *Revistas culturales 2.0*……40

Julia Nantke: Paradigmen digitaler (Re-)Präsentationen von Literatur zwischen Individualisierung und Standardisierung……58

Michael Krewet, Philipp Hegel, Germaine Götzelmann, Danah Tonne, Sybille Söring: Aristoteles aus Reisen: Handschriftenforschung in der digitalen Infrastruktur des SFB 980 „Episteme in Bewegung“……77

Bernhard Runzheimer: „Das ist nicht ganz trivial …“. Die Anpassung gewachsener Projektstrukturen an moderne IT-Standards am Beispiel des *Handschriftencensus*……88

III. Veränderungen in der digitalen Verbundforschung

Mark Fichtner, Tobias Gradl, Canan Hastik: Vom Wandel zur objektbasierten Forschung im Kontext von DARIAH-DE……99

Swantje Dogunke, Timo Steyer: Virtuell Zusammenwachsen: Konzeption, Aufbau und Intention der digitalen Forschungsinfrastruktur im Forschungsverbund MWW……111

Die Beiträge wurden im September 2018 anlässlich des Symposiums „Forschungsinfrastrukturen in den digitalen Geisteswissenschaften. Wie verändern digitale Infrastrukturen die Praxis der Geisteswissenschaften?“ im Rahmen der DFG-geförderten Veranstaltungsreihe „Digitalität in den Geisteswissenschaften“ (http://digitalitaet-geisteswissenschaften.de) präsentiert. Wir danken der Deutschen Forschungsgemeinschaft für die Förderung dieser Tagung.

Einleitung

Martin Huber (Bayreuth)

Seit der Antike sind Bibliotheken, Archive, institutionelle und private Sammlungen bis hin zum Notizenkonvolut die Basis aller wissenschaftlichen Forschung in den Geisteswissenschaften. Schon immer umfassen Forschungsinfrastrukturen in den Geisteswissenschaften gleichermaßen individuelle und institutionelle Sammlungen von Forschungsdaten. Forschungsinfrastrukturen sind seit jeher mit Institutionen wie Akademien oder Universitäten verbunden und repräsentieren deren jeweiliges Konzept von Forschung, Bildung und Wissenschaft. Was aber bedeutet es für diese Institutionen und die mit ihnen verbundenen Fächer der Geisteswissenschaften, wenn als Folge des gegenwärtigen technischen und gesellschaftlichen Wandels Infrastrukturen auf allen Ebenen zunehmend digitale Daten enthalten – oder gar nur noch aus digitalen Daten bestehen? Wie verändern digitale Infrastrukturen die Praxis der Geisteswissenschaften – die Praxis der einzelnen Forschenden, das Selbstverständnis der geisteswissenschaftlichen Disziplinen wie auch die der Institutionen?

Zunächst gibt es für diese Veränderungen, die wir in den Forschungsinfrastrukturen beobachten, keine eindimensionalen Kausalitäten, man wird von einem komplexen Wechselverhältnis ausgehen müssen. In den Disziplinen erweitern sich durch die Möglichkeit mit digitalen Daten zu forschen, die Fragestellungen, es entstehen neue Forschungsgegenstände und Forschungsverfahren. Zugleich wandeln sich die Anforderungen der digital forschenden geisteswissenschaftlichen Disziplinen an ihre Forschungsinfrastrukturen. Da die Transformation der Forschungsdaten im besten Falle ‚wissenschaftsgeleitet' – also im permanenten Austausch mit den Anforderungen der Disziplinen verläuft, beeinflussen digitale Infrastrukturen im Umkehrschluss wiederum selbst in hohem Maße ihrerseits die Disziplinen, da sie als Spiegel des disziplinären Selbstverständnisses und der Praktiken in Forschung und Lehre der Geisteswissenschaften gelten können. An ihren Infrastrukturen lässt sich die Transformation der Geisteswissenschaften durch digitale Ressourcen und Methoden ablesen.

Im Bewusstsein dieser doppelten Perspektive und komplexen Verschränkung von Forschungsinfrastrukturen und den geisteswissenschaftlichen Disziplinen müssen die Fächer sich gegenwärtig darüber verständigen, was alles zur „Infrastruktur" in den digitalen Geisteswissenschaften gehören soll und was ihre digitalen Forschungsdaten charakterisiert, wie sie mit Metadaten angereichert werden sollen und wo sie künftig aufbewahrt werden. Radikal beschleunigt wurde dieser Prozess durch die *Bund-Länder-Vereinbarung zum Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur* (NFDI) vom 26. November 2018, in deren Rahmen in Konkurrenz aller Wissenschaften etwa 30 fachspezifische NFDI-Konsortien in den nächsten 10 Jahren aufgebaut werden. Gegenwärtig sind Antragsverfahren für drei geisteswissenschaftliche NFDI-Konsortien (Text+, Memory und Culture) kurz vor dem Abschluss.

HEIKE NEUROTH und ULRIKE WUTTKE skizzieren vor diesem aktuellen Anlass das große Potential der Geisteswissenschaften, die auf seit Jahrhunderten funktionierende Gedächtnisinstitutionen zurückgreifen können, für den Aufbau von nationalen Forschungsdateninfrastrukturen und formulieren Eckpunkte einer *roadmap* für den Aufbau einer Forschungsdateninfrastruktur Digitale Geisteswissenschaften. Neuroth und Wuttke markieren im Blick auf

die Konkurrenten um die Fördergelder (wie etwa die Astrophysik oder die Klimaforschung) aber auch zentrale Schwächen der Geisteswissenschaften in diesem Transformationsprozess: neben fehlender nationaler (Metadaten)Standards, hemmt vor allem eine noch wenig ausgeprägte Kultur der fachübergreifenden wissenschaftlichen Zusammenarbeit und des Teilens von Forschungsdaten den Aufbau digitaler Infrastrukturen in den Geisteswissenschaften.

Der Paradigmenwechsel den Heike Neuroth und Ulrike Wuttke in der wissenschaftlichen Zusammenarbeit einfordern, steht auch im Zentrum des Beitrags von GERHARD LAUER, in dem er nach den Veränderungen im Gegenstandsfeld, bei den Methoden und den Rollen und Akteuren in der geisteswissenschaftlichen Forschung und Lehre fragt. Eine zentrale Veränderung der Geisteswissenschaften durch digitale Methoden besteht in der Verschiebung des Interesses vom (kanonischen) Einzelwerk hin zum Korpus, oder statistisch gesprochen zur „Grundgesamtheit" eines Gegenstands. Für diese Forschung, für den Aufbau der dafür nötigen digitalen Korpora und die nachhaltige Datenpflege sind neue Berufsfelder und Karrierewege nötig, die auch in den Geisteswissenschaften das Forschen im Team etablieren und formale Modellierungen von geisteswissenschaftlichen Fragestellungen gleichberechtigt neben subjektiven Interpretationen als Kern des Fächer versteht und nachhaltig in der Lehre verankert.

Welchen Einfluss haben digitale Infrastrukturen auf die Episteme der Disziplinen?

HANNO EHRLICHER und JÖRG LEHMANN zeigen am Beispiel einer virtuellen Forschungsumgebung zu spanischsprachigen Kulturzeitschriften den engen Zusammenhang von Datenerhebung und disziplinärer Episteme. In der Forschungsumgebung *Revistas culturales 2.0* wird die Logik, dass die Daten dem epistemischen Prozess vorausgehen umgedreht und die Datenerhebung selbst zu einem substantiellen Teil des epistemischen Prozesses gemacht. Die Erhebung der Metadaten gerät zur epistemischen Herausforderung, da über die Klassifikation ein Höchstmaß an Vernetzung zwischen den einzelnen Zeitschriften bei einem Minimum an Einschränkung für künftige Forschungsinteressen erreicht werden soll. Ehrlicher und Lehmann machen zugleich auf die präformierende Wirkung von Forschungsumgebungen samt Werkzeugen, bereitgestellten Objekten und Metadaten auf die Forscherindividuen aufmerksam und plädieren für eine kritische Reflexion dieser Austauschprozesse gerade auch innerhalb der institutionellen Infrastrukturen.

Digitale Sammlungen generieren nicht nur neue Forschungsfragen, auch die Forschungsgegenstände erhalten durch digitale Darstellungsformen eine veränderte Semantik. Im Blick auf einen repräsentativen Querschnitt unterschiedlicher digitaler Plattformen der Literaturwissenschaft – von werk- oder autorzentrierten Portalen, über digitale Editionen bis hin zu unspezifischen digitalen Textsammlungen – zeigt JULIA NANTKE die Besonderheiten der digitalen Medialität, in der die literarischen Texten (re)präsentiert werden. Sie konstatiert im Spannungsfeld von Plurarisierung und Standardisierung eine Entautomatisierung der Repräsentationsstrukturen bei den digitalen Präsentationsformen und entdeckt zugleich eine neue Semantisierung der Literatur. Sie entsteht durch die Multiplikation der Perspektiven auf die Gegenstände, die durch die mediale Zusammenstellung und Einbettung in Kontexte in Antizipation von Nutzungsszenarien. Insbesondere die in digitalen Präsentationsformen er-

wünschte Usability durch Visualisierung der Literatur, erzeugt neue Fragen an die Materialität der Literatur.

Welche Rolle spielen Forschungsinfrastrukturen derzeit für die Forschung in den geisteswissenschaftlichen Fächern?

Die Transformation individuell entwickelter und über einen längeren Zeitraum mit der fallweisen Förderung gewachsener, vom Fach dringend benötigter und gut genutzter Datenbanken in institutionelle Infrastrukturen ist keineswegs trivial. BERNHARD RUNZHEIMER führt am Beispiel des Marburger *Handschriftencensus* zur Erfassung des deutschsprachigen Handschriftenerbes des Mittelalters vor, welche komplexen technischen Prozesse und auch welche Verständnisschwierigkeiten zwischen der wissenschaftlichen und der informationstechnischen Seite bei einem digitalen Infrastrukturprojekt anfallen können.
Digitale Infrastrukturen können nicht nur neue Forschungsfragen generieren, sie verändern auch das Forschen im Fach und beeinflussen die alltägliche Forschungspraxis der betreffenden Forscherinnen und Forscher – vor allem in der Verbundforschung.

MICHAEL KREWET, PHILIPP HEGEL, GERMAINE GÖTZELMANN, DANAH TONNE und SYBILLE SÖRING beschreiben die Veränderungen, die an der digitale Forschungsinfrastruktur des SFB „Episteme in Bewegung“ für die Handschriftenforschung in der Philosophie zu beobachten sind am Beispiel von Aristoteles’ *de interpretatione*, einer antiken Schrift, die in ca. 150 Handschriften und zahllosen annotierten Kopien erhalten ist. Neben der erwartbaren Möglichkeit, qualitative *und* quantitative Analysen vorzunehmen, die durch das Nebeneinander von digitalen Scans und den Kollationen und Transkriptionen der Handschriften möglich sind, werden positive Effekte des digitalen Datenrepositoriums genannt, die überraschen: auch kleinteilige Ergebnisse, die für eine analoge Veröffentlichung zu wenig an kritischer Masse bieten, und erst im Verbund mit anderem relevant für eine Neuinterpretation werden, gehen jetzt nicht mehr verloren, sondern bleiben nachhaltig verfügbar. Durch die direkte Integration des Repositoriums in den Alltag der Forschenden wird zudem kollaboratives Arbeiten auch über unterschiedliche Standorte hinweg begünstigt. Gleichzeitig macht dieser Beitrag auf ein zentrales Abstimmungsproblem beim Aufbau übergreifender, nationaler Forschungsdateninfrastrukturen aufmerksam. Es wird die Balance gefunden werden müssen, zwischen der Notwendigkeit Normdaten und Taxonomien anzubieten und dem berechtigten Wunsch der Disziplinen nach projektspezifisch modifizierten Ontologien. Denn die Metadaten sind epistemische Entscheidungen jeder Disziplin.

Diversität der Forschungsdaten und ein damit einhergehendes breite Spektrum an Metadatenformaten sind ein Spezifikum großer verteilter Forschungsdatensammlungen, wie etwa DARIAH-DE. Der Beitrag von MARK FICHTNER, TOBIAS GRADL und CANAN HASTIK beschreibt wie sich die Datenföderationsarchitektur in DARIAH durch die Integration von Objektsammlungen aus Museen und der Kunstwissenschaft wandelt. Im Gegensatz zur dokumentbasierten Datenhaltung, in der versucht wird, alle Informationen im Dokument zu „kapseln“, wird bei der objektbasierten Datenhaltung der Kontext (Personendaten, Ortsinformation, Zeitangaben) meist separat erfasst. Das Ziel für DARIAH ist eine einheitliche Datensammlung, die die Heterogenität der Metadaten weitgehend integriert, um eine vernetzte, zunehmend disziplinübergreifende Forschung in den Geistes- Kultur und Kunstwissenschaften zu ermöglichen.

Eine gemeinsame Forschungsinfrastruktur dreier herausragender Gedächtnisinstitutionen wie das Deutsche Literaturarchiv Marbach, die Klassik Stiftung Weimar und die Herzog August Bibliothek Wolfenbüttel steht im Forschungsverbund MWW vor der Aufgabe über die digitale Infrastruktur einen Mehrwert zu schaffen. SWANTJE DOGUNKE und TIMO STEYER geben in ihrem Werkstattbericht einen Einblick in die Entwicklungsarbeit der letzten fünf Jahre, die das Ziel verfolgt, aus drei Institutionen mit je eigenen Nutzungsprofilen und Forschungsprojekten eine gemeinsame Community zu formen. Ein wichtiges Ergebnis dieses Prozesses ist neben einer engen Kommunikation mit den Nutzerinnen und Nutzern hinsichtlich deren Anforderungen und Wünschen auch ein Bewusstsein dafür, dass die Infrastruktur mehr ist als Dienstleistung, die über das reine Anbieten von Daten und das Bereitstellen von Tools hinausgeht und auch eine eigene Forschungsleistung darstellen kann.

***

Infrastrukturen sind komplexe institutionelle Gefüge, deren digitale Transformation nur in einem permanenten Austausch der beteiligten Akteure – also in der Kommunikation zwischen wissenschaftlichen Forscherinnen und Forschern, den Fachcommunities und den Gedächtnisinstitutionen gelingen wird. An allen Einzelbeispielen ist deutlich geworden, dass die Fächer in den Geisteswissenschaften gut daran täten, kollaboratives Arbeiten mit gemeinsamen Daten und in den gleichen Datenräumen als einen gewichtigen Teil ihrer Ausbildung künftiger Forscherinnen und Forscher zu verstehen und als Forschungsleistung ebenso hoch zu bewerten, wie die Forschung von Einzelnen. Die Annotierung von Daten, das Erstellen von digitalen Ressourcen für das eigene Fach ist eine wissenschaftliche Leistung und muss auch als eine solche anerkannt werden, die Karrieren begründen kann. Nehmen wir diese Tätigkeiten ernst, dann entstehen auch veränderte Berufsbilder in den Geisteswissenschaften. Nehmen wir die digitale Transformation der Geisteswissenschaften ernst, dann sollten wir als Geisteswissenschaften eine Kritik der Datensätze etablieren und Orte für Rezensionen zu digitalen Ressourcen schaffen. So werden auch digitale Forschungsprojekte genuin aus disziplinären Fragestellungen und weniger von den technischen Möglichkeiten her entwickelt.

I. Grundlagen

Eine gemeinsame geisteswissenschaftliche Forschungsdatenkultur: Utopie oder Dystopie?

Heike Neuroth, Ulrike Wuttke (Potsdam)

1. Wissenschafts- und förderpolitischer Rahmen

1.1 Einleitung

Durch die digitale Transformation der Forschungsprozesse ist ein starker Anstieg der Produktion, Aggregation und Nachnutzung von digitalen Forschungsdaten in allen Wissenschaftsdisziplinen zu verzeichnen.[1] Die digitale Transformation und die damit zusammenhänge Datenflut (*data deluge*) betrifft nicht nur die Wissenschaft, sondern auch die Wirtschaft, die Politik, die Verwaltung und die breitere Gesellschaft. Im Rahmen der förderpolitischen Debatte ist die sachgemäße Aufbewahrung und Bereitstellung von Forschungsdaten in Deutschland zum Beispiel in die *Vorschläge zur Sicherung guter wissenschaftlicher Praxis*[2] der DFG aufgenommen bzw. wird in den *Leitlinien zum Umgang mit Forschungsdaten*[3] der DFG erläutert. Einzelne Fachdisziplinen haben nach dem Vorbild der DFG-Richtlinien eigene fachspezifische Empfehlungen veröffentlicht, darunter z. B. die Biodiversität, Soziologie, Sprachkorpora etc.[4] Neben dem Argument, dass die Ergebnisse öffentlich geförderter Forschung auch der breiten Öffentlichkeit zur Verfügung stehen sollten, werden diesbezüglich u. a. die Sicherung der wissenschaftlichen Qualität und der Nachvollziehbarkeit der Forschung anhand der unmittelbaren Datengrundlage als Argumente angeführt. Gerade in Zeiten von *fake journals* und dem Zurückziehen von Artikeln spielt die Prüfung der Datengrundlage eine zunehmende Rolle. So hat z. B. die Fachzeitschrift *JAMA* 2018 dreizehn Artikel des weltweit anerkannten Ernährungspsychologen Brian Wansink (Cornell University) zurückgezogen, dessen Studien 20.000mal zitiert wurden, weil die Ergebnisse nicht validierbar sind, u. a. auch, weil kein Zugriff auf die Originaldaten möglich war.[5]

2017 hat sich die Initiative *Collections as Data* in den USA gegründet, die von dem Institute of Museum and Library Services initiiert wurde. Hier wurden 10 Prinzipien formuliert, die quasi parallel zu anderen Initiativen (s.u.) inhaltlich und strategisch in die gleiche Richtung gehen. So lautet zum Beispiel ein Prinzip: „Collections as data development aims to encourage computational use of digitized and born digital collections“ oder: „Collections as Data should be made openly accessible by default, except in cases where ethical or legal obligations preclude it.“ Allein diese beiden Prinzipien zielen darauf ab, dass die Sammlungen aus den sogenannten

[1] vgl. Borgman, Christine L.: Big Data, Little Data, No Data. Scholarship in the Networked World. Cambridge (Mass.), London: MIT Press 2015, S. 3-16.

[2] DFG - Deutsche Forschungsgemeinschaft: Vorschläge zur Sicherung guter wissenschaftlicher Praxis / Proposals for Safeguarding Good Scientific Pratice, Denkschrift / Memorandum, Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“ / Recommendations of the Commission on Professional Self Regulation in Science. Bonn: Wiley-VCH 2013, S. 21-22.

[3] DFG - Deutsche Forschungsgemeinschaft: Leitlinien zum Umgang mit Forschungsdaten. 2015. https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf.

[4] vgl. DFG - Deutsche Forschungsgemeinschaft: Umgang mit Forschungsdaten. https://www.dfg.de/foerderung/antrag_gutachter_gremien/antragstellende/nachnutzung_forschungsdaten/.

[5] vgl. z. B. Lingenhöhl, Daniel: Der tiefe Fall eines Ernährungspapstes. Spektrum.de News 06.03.2019. https://www.spektrum.de/news/der-tiefe-fall-eines-ernaehrungspapstes/1593004 .

Gedächnisinstitutionen wie Bibliotheken, Museen und Archive als Daten-Grundlage für digitale (automatische) Verfahren und Methoden dienen und ohne weitere Mühen für die Nachnutzung zur Verfügung stehen sollten.

Im Zusammenhang mit der allgemeinen Forderung nach der Bereitstellung von Forschungsdaten als Open Research Data (ORD) im Sinne der Open Science-Bewegung, steht in jüngster Zeit verstärkt ihre Bereitstellung gemäß den sogenannten FAIR-Prinzipien[6] im Vordergrund und rückt somit die Eignung der Daten für die Nachnutzung in den Mittelpunkt des Interesses. FAIR steht für „Findable, Accessible, Interaperable, Reuseable" und gilt für die intellektuelle Nachnutzung durch Menschen, aber besonders für die maschinelle Nachnutzung (*machine-actionability*), d.h. „the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention"[7]. Die Europäische Kommission hat im November im Rahmen der EOSC (European Open Science Cloud) u. v. a. a. den Bericht *Turning FAIR into reality*[8] veröffentlicht, der die breite Spanne an notwendigen Änderungen beschreibt, um die sogenannte FAIRness von Daten sicherzustellen.

Durch diese Entwicklungen werden Forschungsdatenmanagement und aktive Forschungsdatenmanagementpläne (DMPs)[9] immer wichtiger und auch die Geisteswissenschaften stehen nun vor der Herausforderung, ein grundlegendes Verständnis dafür zu entwickeln:

- was geisteswissenschaftliche Forschungsdaten sind,
- wie ihr Umgang systematisch geregelt werden kann,
- welche Kernkompetenzen in der Lehre bzw. in der Weiterbildung vermittelt werden sollten und
- welche systemischen Veränderungen (Stichworte Förder- und Evaluationsprozesse) notwendig sind, um die zunehmende Digitalisierung der Geisteswissenschaften (Stichwort Digital Humanities) zu unterstützen.

Hierbei gilt es zunächst, den Forschungsdatendiskurs in den Geisteswissenschaften überdisziplinär zu öffnen, um einen allgemeinen definitorischen Konsens zu erreichen, der auch Erfahrungen aus bereits etablierten Forschungsdatenkulturen einbezieht. Erst dann kann eine Verfeinerung von disziplinspezifischen Fragestellungen vorgenommen werden. Im Folgenden werden Handlungsbedarfe und mögliche Lösungsszenarien bei der Etablierung einer gemeinsamen Forschungsdatenkultur in den Geisteswissenschaften vor dem Hintergrund aktueller Entwicklungen und grundsätzlicher Prämissen skizziert.

1.2 Aktuelle Entwicklungen

Neben den bereits oben genannten europäischen Veröffentlichungen sind in letzter Zeit eine Reihe wichtiger Veröffentlichungen auf nationaler Ebene erschienen, die einen großen Einfluss auf die Praktiken der Wissenschaftler*innen und die förderpolitischen Rahmenbedingungen haben werden. So hat der Deutsche Bundestag im September 2018 *Die Hightech-Strategie 2025 – Forschung und Innovation für die Menschen* veröffentlicht, in der es sehr konkret heißt:

[6] vgl. Wilkinson, Mark D. et al.: The FAIR Guiding Principles for Scientific Data Management and Stewardship. in: Scientific Data 3 2016.
[7] GO FAIR: FAIR Principles. https://www.go-fair.org/fair-principles/.
[8] European Commission Expert Group on FAIR Data: Turning FAIR into Reality. Final Report and Action Plan from the European Commission Expert Group on FAIR Data. European Commission: Brüssel 2018.
[9] vgl. z. B. Neuroth, Heike / Engelhardt, Claudia / Klar, Jochen / Enke, Harry: Aktives Forschungsdatenmanagement - Research Data Management Organiser. In: ABI Technik 38,1 (2018), S. 55-64.

- Die umfassende Nutzung von Forschungsdaten ermöglicht die Beantwortung neuer, auch multidisziplinärer wissenschaftlicher Fragen.
- Für die Nutzung von Daten müssen international anschlussfähige Lösungen gefunden werden. Um unsere Wettbewerbsfähigkeit in einer datengetriebenen Wissenschaft und Wirtschaft zu sichern, ergänzen wir unsere Förderung durch übergreifende Infrastruktur-Ansätze wie den Aufbau einer nationalen Forschungsdateninfrastruktur sowie einer European Open Science Cloud.[10]

Konsequenterweise folgte im November 2018 die *Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI)*, die in ihrer Präambel auf die Notwendigkeit von potentiellen NFDI-Konsortien als „wissenschaftlich breit nutzbare Datenschätze mit gesellschaftlichem Mehrwert“[11] hinweist. In dem zugrunde liegenden Förderprogramm wird darauf hingewiesen, dass „ein stimmiges Konzept zu Datennutzung und -zugang sowie Auffindbarkeit und Nachnutzbarkeit der Daten, welches entlang der FAIR-Prinzipien ausgerichtet ist“[12], vorliegen muss. Dafür stehen in den nächsten zehn Jahren bis zu 90 Millionen Euro pro Jahr im Endausbau für insgesamt ca. 30 sogenannte fachspezifische NFDI-Konsortien zur Verfügung. Im Rahmen der RDA-DE Tagung im Februar 2019 am GFZ Potsdam[13] haben sich bereits einige potenzielle fachspezifische NFDI-Konsortien vorgestellt[14], die in den Geisteswissenschaften verankert sind (u. a. NFDI4Text+, NFDI4Memory und NFDI4Culture). Anfang 2019 hat sich zudem das Geisteswissenschaftliche Forum NDFI für den Informationsaustausch gegründet, an dem sich bereits über 20 geisteswissenschaftliche Fachverbände bzw. Fachgesellschaften beteiligen.

2. Geisteswissenschaftlicher Kontext

Die bisherige Forschungsdatendiskussion in den Geisteswissenschaften ist von zwei Tendenzen geprägt: der besonderen Spezifik geisteswissenschaftlicher Forschungsdaten[15] sowie der Notwendigkeit der Abgrenzung einzelner Disziplinen. Zum einen haben Geisteswissenschaftler*innen ein eher ambivalentes Verhältnis zum Datenbegriff.[16] Zum anderen sind auf Grund der disziplinären Vielfalt innerhalb der Geisteswissenschaften und weil potenziell eine breite Spanne von Materialien als geisteswissenschaftlicher Untersuchungsgegenstand in Frage

[10] BMBF - Bundesministerium für Bildung und Forschung: Forschung und Innovation für die Menschen: Die Hightech-Strategie 2025. Berlin 2018, S. 55.
[11] GWK - Gemeinsame Wissenschaftskonferenz: Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI) vom 26. November 2018 (BAnz AT 21.12.2018 B10), S. 1.
[12] GWK - Gemeinsame Wissenschaftskonferenz: Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI) vom 26. November 2018 (BAnz AT 21.12.2018 B10), Paragraph 5, Abschnitt g.
[13] vgl. RDA DE: RDA Deutschland Tagung 2019. https://www.rda-deutschland.de/events/tagung-2019.
[14] Einige haben dazu die Möglichkeit einer Kurzvorstellung im Plenum genutzt, andere das Posterformat gewählt.
[15] vgl. u. a. Sahle, Patrick / Kronenwett, Simone: Jenseits der Daten. Überlegungen zu Datenzentren für die Geisteswissenschaften am Beispiel des Kölner ‚Data Center for the Humanities'. In: LIBREAS. Library Ideas, 23 (2013), S. 76–96, Buddenbohm, Stefan / Engelhardt, Claudia / Wuttke, Ulrike: Angebotsgenese für ein geisteswissenschaftliches Forschungsdatenzentrum. In: Zeitschrift für digitale Geisteswissenschaften 2016,1. Open Research Data Task Force: Realising the Potential. Final Report of the Open Research Data Task Force“. Open Research Data Task Force, 2018, S. 6.
[16] vgl. u. a. Cremer, Fabian / Klaffki, Lisa / Steyer, Timo: Der Chimäre auf der Spur: Forschungsdaten in den Geisteswissenschaften. In: O-Bib 5,2 (2018), S. 142-162.

kommt, geisteswissenschaftliche Forschungsdaten und die für ihre Interpretation benötigte Expertise sehr heterogen. Die Natur geisteswissenschaftlicher Forschungsdaten kann berechtigterweise nicht nur gestellt werden als „Was sind geisteswissenschaftliche Daten?“ sondern auch als „Wann sind es geisteswissenschaftliche Daten?“:

> The humanities are in a position similar to the social sciences in determining what or when something might be used as data. […] Several characteristics of the humanities distinguish them from the sciences and social sciences with respect to data scholarship. One is their focus on interpretation and reinterpretation. [...] A second characteristic is the vast variety of data sources and the expertise to interpret them. [...] As a result of these two qualities, data sources can be reused repeatedly, often becoming more valuable as they cumulate over time. Conversely, because the same objects can be represented and interpreted in so many ways, it is difficult to apply classification mechanisms that transcend the array of possible uses of collections.[17]

Geisteswissenschaftliche Forschungsdaten können sowohl analoger als auch digitaler Natur sein und in Form von Texten, Bildern, Audio-, Multimedia- oder 3D-Daten vorliegen. Diese Daten können darüber hinaus in verschiedenen (Meta-)Datenformaten in unterschiedlichen Qualitätsstufen, z. B. im Fall von mittels OCR erstellten Volltexten oder Transkriptionen im Douple-Keying-Verfahren, vorliegen. In vielen Fällen liegen die (Meta-)Daten mehrsprachig und zum Teil in nicht-lateinischen Schriften vor. Oftmals sind die Daten sehr stark untereinander vernetzt, da sie mehrere Kommentar- bzw. Interpretationsschichten reflektieren. Gerade der letzte Punkt, die Subjektivität geisteswissenschaftlicher Forschung und der daraus resultierenden Forschungsdaten, stellt eine weitere Herausforderung dar.[18]

Im Rahmen von DARIAH-DE (Digitale Forschungsinfrastruktur für die Geistes- und Kulturwissenschaften) wurde sich auf die folgende Definition für Forschungsdaten geeinigt:

> Unter digitalen geistes- und kulturwissenschaftlichen Forschungsdaten werden innerhalb von DARIAH-DE all jene Quellen/Materialien und Ergebnisse verstanden, die im Kontext einer geistes- und kulturwissenschaftlichen Forschungsfrage gesammelt, erzeugt, beschrieben und/oder ausgewertet werden und in maschinenlesbarer Form zum Zwecke der Archivierung, Zitierbarkeit und zur weiteren Verarbeitung aufbewahrt werden können.[19]

Die oben angerissenen epistemologischen Fragestellungen zur Natur geisteswissenschaftlicher Daten können in diesem Rahmen nicht im Detail ausgeführt werden. Es soll jedoch festgehalten werden, dass geisteswissenschaftliche Forschungsdaten, im Gegensatz zu vielen naturwissenschaftlichen Forschungsdaten, in der Regel besonders langlebig sind. Ein historisches Korpus z. B. in Form einer Datenbank oder einer digitalen Edition, Vorhaben, an denen oft lange Jahre gearbeitet wurden, ist für lange Zeit eine autoritative Quelle. Dazu kommt, dass es trotz gewisser Gemeinsamkeiten, die im Folgenden eruiert werden, trügerisch ist, von DEN Geisteswissenschaften oder DEN Digital Humanities zu sprechen, da sich hinter diesen

[17] Borgman 2015, S. 166.

[18] vgl. u. a. Raspe, Martin / Schelbert, Georg: Genau, wahrscheinlich, eher nicht. Beziehungsprobleme in einem kunsthistorischen Wissensgraph. In: Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten. Hg. von Andreas Kuczera / Thorsten Wübbena / Thomas Kollatz. 2019 (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 4).

[19] DARIAH-DE: Weiterführende Informationen. https://de.dariah.eu/weiterfuhrende-informationen.

Begriffen sehr unterschiedliche Daten, Methoden und Verfahren, Geschwindigkeiten der Digitalisierung sowie Anforderungen und Bedarfe verbergen.

Auch außerhalb der Geisteswissenschaften gibt es bisher keine standardisierte Definition von Forschungsdaten in Deutschland. Alle bisherigen Beschreibungen wurden bewusst offen formuliert.[20] In einigen naturwissenschaftlichen Fächern (wie z. B. der Astrophysik[21] oder der Klimaforschung[22]) wurde jedoch bereits ein Konsens darüber erreicht, was Forschungsdaten sind und wie sie beschrieben und aufbewahrt werden. Ein solcher Lösungsansatz (*Good Practice*) aus einer etablierten Forschungsdatenkultur kommt aus der Astrophysik. Da in dieser Fachdisziplin die benötigten Großgeräte sehr kostspielig sind, beteiligen sich in der Regel zahlreiche Länder an ihrer Entwicklung sowie den Investitions- und Betriebskosten. Dies führt dazu, dass eine Kultur der internationalen Zusammenarbeit entwickelt wurde bzw. werden musste, d. h. die Wissenschaftler*innen teilen sich die Beobachtungszeit und in der Regel auch die Forschungsdaten. Hieraus entwickelte sich eine gemeinsame Forschungsdatenkultur, die zu einer Reihe weiterer Standardisierungen bezüglich Werkzeuge, Dienste, Datenformate etc. geführt hat. Als ein Beispiel sei hier der *Sloan Digital Sky Survey* (SDSS) angeführt, der im Jahr 2000 den Startschuss für die „modernste Kartierung des Himmels“[23] eröffnete. Mit der mittlerweile 15. Datenpublikation (*data release*) bildeten diese Daten bisher die Grundlage für mehr als 8.000 wissenschaftliche Publikationen und es entstand ein Katalog von mehr als 200 Millionen Galaxien sowie Millionen von Sternen und Quasaren. In einem Hinweis zur Datenzitierung heißt es: „If you have used public SDSS data in your paper, please cite the following papers describing the instruments, survey, and data analysis as appropriate.“[24]

In den Geisteswissenschaften fehlt dagegen weitestgehend ein Konsens darüber, was Forschungsdaten sind und vor allem, wie sie beschrieben und aufbewahrt werden sollten, sowohl in den einzelnen Fachdisziplinen als auch darüber hinaus. So produziert das Deutsche Textarchiv (DTA) sehr viele Forschungsdaten(-sammlungen), z. B. hochauflösende Digitalisate, Transkriptionen als Volltexte, Annotationen, TEI-Auszeichnungen, zahlreiche Exportformate und unterschiedliche Metadaten sowie Dokumentationen. Was davon sind Forschungsdaten? Was muss wie aufbewahrt werden? Welche, auch zukünftigen, Nutzungsszenarien sind vorstellbar? Ist eine Unterscheidung zwischen aktiver und passiver Nutzung von Forschungsdaten sinnvoll, weil manche Forschungsdatensammlungen eher den Anspruch haben, nur langzeitarchiviert zu werden, um der guten wissenschaftlichen Praxis zu dienen?[25] Welche Metadatenstandards sind verbindlich und international interoperabel? Welcher Status kommt Forschungssoftware zu? Wem gehören die Forschungsdaten in den Geisteswissenschaften?

---

[20] So z. B. die Definition der Arbeitsgruppe Forschungsdaten der Allianz der deutschen Wissenschaftsorganisationen (vgl. Schwerpunktinitiative „Digitale Informationen“: Forschungsdaten. https://www.allianzinitiative.de/archiv/forschungsdaten/).

[21] vgl. z. B. NASA Science: Astrophysics Data Centers. https://science.nasa.gov/astrophysics/astrophysics-data-centers.

[22] vgl. z. B. NOAA - National Oceanic and Atmospheric Administration - Earth System Research Laboratory (ESRL) - Physical Sciences Division (PSD): Data and Imagery. https://www.esrl.noaa.gov/psd/data/.

[23] Max-Planck-Institut für Astronomie: Startschuss für modernste Kartierung des Himmels. Webseite.

[24] SDSS - Sloan Digital Sky Survey. How to Cite SDSS. https://www.sdss.org/collaboration/citing-sdss/.

[25] vgl. Pronk, Tessa E.: The Time Efficiency Gain in Sharing and Reuse of Research Data, In: Data Science Journal 18,1 (2019), S. 10: „Moreover, results suggest overall efficiency can be increased if not too much effort is put into sharing datasets with low probability of reuse.” Vgl. auch DFG – Deutsche Forschungsgemeinschaft: Leitlinien zum Umgang mit Forschungsdaten. 2015, S. 1 „Den Regeln der Guten Wissenschaftlichen Praxis folgend sollen Forschungsdaten in der eigenen Einrichtung oder in einer fachlich einschlägigen, überregionalen Infrastruktur für mindestens 10 Jahre archiviert werden.”

Hinzu kommt, dass auch organisatorische und strukturelle Fragen in den Geisteswissenschaften noch unzureichend geklärt sind. Während zahlreiche naturwissenschaftliche Disziplinen, großteils in engen internationalen Kooperationen, eigene Forschungsdatenzentren aufgebaut haben, stehen ähnliche Bestrebungen in den Geisteswissenschaften noch am Anfang. Die im Vergleich sehr junge Klimaforschung hat es zum Beispiel sehr früh verstanden, sowohl in Deutschland als auch auf internationaler Ebene, eng zu kooperieren und gemeinsame Standards und Werkzeuge zu entwickeln. Obwohl hierfür alleine in Deutschland die Zusammenarbeit zwischen mehr als 100 universitären und außeruniversitären Forschungseinrichtungen organisiert werden musste, steht mit verschiedenen Förderinstrumenten und zahlreichen Fachgremien eine tragfähige Struktur zur Verfügung, die das breite und interdisziplinäre Feld der Klimaforschung (z. B. Meteorologie, Ozeanographie, Politologie, Ökonomie, Soziologie etc.) organisieren hilft. Mit dem Deutschen Klimarechenzentrum (DKRZ) existiert in Deutschland auch ein für die Archivierung und Zitierung der Forschungsdaten verantwortlicher Partner.

In Analogie hierzu ist in den Geisteswissenschaften vieles noch unklar bzw. nicht geregelt. Welche Rolle kommt Gedächtnisinstitutionen wie Bibliotheken, Archiven und Museen zu, die eine große Menge potenzieller geisteswissenschaftlicher Forschungsdaten produzieren, in dem sie ihre Sammlungen digitalisieren und zur Verfügung stellen? Wie sollten einheitliche Regelungen für die Nachnutzungsbedingungen für diese digitalen Sammlungen aussehen, so dass Wissenschaftler*innen rechtssicher darauf zugreifen und diese nachnutzen können? Allein der momentan geringe und wenig systematische Stand der Digitalisierung von geisteswissenschaftlichen Quellen bzw. Sammlungen in den Gedächtnisinstitutionen stellt ein Problem für datengetriebene Forschungsfragen dar. Mangels genauen Berechnungen sei hier auf die Schätzung von Michael Knoche verwiesen, laut der „erst ein niedriger zweistelliger Prozentsatz der nationalen Buchproduktion in elektronischer Form zur Verfügung“ steht, weil Ressourcen zur Digitalisierung fehlen und das restriktive deutsche Urheberrecht eine erhebliche Barriere darstellt.[26] Hinzu kommt die Heterogenität der Beschreibungsformate (Metadaten), der Dateiformate, Auszeichnungssprachen etc., die maschinelle Auswertungsverfahrenen über mehrere digitale Sammlungen (Korpora) hinweg mit einem vertretbaren Zeitaufwand nahezu unmöglich macht.

Ohne einen disziplinübergreifenden Forschungsdatendiskurs in den Geisteswissenschaften unter Einbeziehung der Gedächtnisinstitutionen wird es kaum möglich sein, auf nationaler Ebene ein gemeinsames Verständnis einer geisteswissenschaftlichen Forschungsdatenkultur zu etablieren. Diese ist jedoch die Voraussetzung:

- für nationale und internationale interdisziplinäre Kooperationen und damit einhergehende Standardisierungen,
- um vor allem zukünftigen Anforderungen der Förderer (bzw. auch Politik und Gesellschaft) im Kontext der FAIR-Prinzipien zu begegnen,
- um eine gemeinsame gute datenwissenschaftliche Praxis zu etablieren,
- um gemeinsam den Herausforderungen der digitalen Transformation zu begegnen mit abgestimmten Notwendigkeiten in Entwicklungen, Förderinstrumenten etc.,

[26] vgl. Knoche, Michael: Die Idee der Bibliothek und ihre Zukunft. Göttingen: Wallstein 2018, S. 67-73, Zitat S. 69. Auch Klaffki, Lisa / Schmunk, Stefan / Stäcker, Thomas: Stand der Kulturgutdigitalisierung in Deutschland: Eine Analyse und Handlungsvorschläge des DARIAH-DE Stakeholdergremiums „Wissenschaftliche Sammlungen“, Göttingen 2018 (DARIAH-DE Working Papers, 26), S. 6-17 verzeichnet einen ähnlich spärlichen Umfang der Kulturgutdigitalisierung in Deutschland.

- um koordiniert (Mindest)Standards, Policies etc. zu entwickeln, die jeweils fachspezifisch ausdifferenziert werden zu können und
- für die Entwicklung der notwendigen Lehr-, Weiterbildungs-, und Schulungskonzepte.

Konsequenterweise heißt es auf der Webseite der Gemeinsamen Wissenschaftskonferenz (GWK) zur NFDI:

> Derzeit oft dezentral, projektförmig und temporär gelagerte Datenbestände von Wissenschaft und Forschung sollen im Rahmen der NFDI für das gesamte deutsche Wissenschaftssystem systematisch erschlossen werden. Die NFDI soll Standards im Datenmanagement setzen und als digitaler, regional verteilter und vernetzter Wissensspeicher Forschungsdaten nachhaltig sichern und nutzbar machen […]. Auf diese Weise wird eine unverzichtbare Voraussetzung dafür geschaffen, neue wissenschaftliche Erkenntnisse zu gewinnen und Innovationen in Forschung und Gesellschaft zu ermöglichen. [27]

3. Potentielle Roadmap Forschungsdateninfrastruktur digitale Geisteswissenschaften

3.1 Gemeinsame Forschungsdatenkultur

Ausgangsbasis für die nachfolgenden Überlegungen stellen die in der *Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI)* vom 26. November 2018 formulierte Ziele dar.[28] Diese betonen sehr stark Aspekte der Gemeinsamkeit, Koordination, Anbindung an nationale sowie internationale Entwicklungen, Zusammenarbeit, Standardisierung sowie fachübergreifende Prozesse und Strukturen. Hieraus ergibt sich die Notwendigkeit für die Geisteswissenschaften, sich gemeinsam auf den (langen) Weg zu begeben und übergeordnete Diskussionen und Lösungswege anzustoßen. Zu kleinteilige Fokussierungen auf die jeweiligen Fachdisziplinen mit ihren spezifischen Forschungsfragen werden in der Zukunft nicht mehr zielführend sein.

Wie könnte eine gemeinsame geisteswissenschaftliche Forschungsdatenkultur aussehen? Ein erster Schritt der Annäherung an diese Frage ist eine grobe Unterscheidung der Entstehungskontexte geisteswissenschaftlicher Forschungsdaten, diese entstehen meist:

- im Kontext von konkreten Forschungsfragen, in den meisten Fällen im Rahmen (kleinerer) Drittmittel-basierter Projekte sowie
- beim Aufbau von digitalen (größeren) Sammlungen, sozusagen als Grundlagenforschung, analog zu den Großgeräten in den Naturwissenschaften und als Basis für die Erforschung breiter, zukünftiger Forschungsfragen.[29]

Allerdings scheinen die Einzelvorhaben oder kleinen Projektverbünde in den Geisteswissenschaften zu dominieren. So sind über die letzten Jahrzehnte hinweg zahlreiche digitale Daten-Silos entstanden, zwischen denen kaum ein Austausch, sei es auf der eigentlichen Datenebene

[27] GWK - Gemeinsame Wissenschaftskonferenz: Informationsinfrastrukturen / NFDI. https://www.gwk-bonn.de/themen/weitere-arbeitsgebiete/informationsinfrastrukturen-nfdi/.
[28] vgl. GWK 2018, Paragraph 1.
[29] vgl. Schöch, Christof: Aufbau von Datensammlungen. In: Jannidis, Fotis / Hubertus Kohle / Malte Rehbein (Hg). Digital Humanities. Eine Einführung. Stuttgart: J.B. Metzler, 2017, S. 223-233. Schöch unterscheidet drei Typen von Datensammlungen (Projektbezogene Datensammlungen, Gemeinschafts-Datensammlungen, Referenz-Datensammlungen).

oder auch nur auf der Ebene der Metadaten, möglich ist. Dazu kommen Nachhaltigkeitsfragen: Wie kann sichergestellt werden, dass diese Daten auch in Zukunft gefunden und zugänglich gemacht werden können, wenn sie in den meisten Fällen auf Projekthomepages liegen (dies betrifft schon den Bereich „Findable“ und „Accessible“ der FAIR-Prinzipien)? Das wichtigste Ziel der direkten Nachnutzbarkeit („Reusable“) wird bisher noch so gut wie gar nicht in den Projektvorhaben adressiert. Hier fehlen Strukturen und Prozesse für die Überführung in vertrauenswürdige Datenzentren.[30] Diesen Entscheidungen müssen natürlich Absprachen über die Aufbereitung der Daten, geeignete Standards, Schnittstellen und Formate vorausgehen, an die sich alle Datenproduzenten halten müssen.

Auch der Bereich des breiter angelegten Aufbaus größerer digitaler Sammlungen ist in Deutschland wenig koordiniert. So gibt es zahlreiche gelungene Einzelbeispiele wie zum Beispiel das DTA als „Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache“[31] oder die Digitale Bibliothek in TextGrid als „umfangreiche Sammlung XML/TEI-erschlossener Texte aus Belletristik und Sachliteratur vom Anfang des Buchdrucks bis zu den ersten Jahrzehnten des 20. Jahrhunderts“[32]. Allerdings gibt es in Deutschland keinen nationalen Ansatz, alle digitalen Sammlungen basierend auf gemeinsamen standardisierten Vorgaben zu digitalisieren, zu erschließen und für die Nachnutzung nachhaltig als digitalen Wissensspeicher zur Verfügung zu stellen. Als Vorbild oder Inspiration kann HathiTrust in den USA dienen, das seit 2008 aufgebaut wird. Als Ziel wird Folgendes formuliert: „provides long-term preservation and access services to digitized content from a variety of sources, including Google, the Internet Archive, Microsoft, and in-house member institution initiatives“[33]. Mit Bezug zu den Zielen der FAIR-Prinzipien, die Maschinenlesbarkeit der (Meta-)daten zu erhöhen, betont das HathiTrust Research Center (HTRC) das Ziel „computational analysis of the HathiTrust corpus“ zu ermöglichen[34] und stellt dafür eine technische Infrastruktur mit einer Reihe von geeigneten digitalen Werkzeugen und Diensten zur Verfügung.

Es ist eine schöne Vorstellung, wenn die digitalen Bestände z. B. der Bayerischen Staatsbibliothek mit dem Deutschen Textarchiv, der Digitalen Bibliothek in TextGrid sowie den im *Zentralen Verzeichnis Digitalisierter Drucke* (ZVDD*)* nachgewiesenen Digitalisaten in einem deutschen Wissensspeicher zugänglich und nachnutzbar wären, versehen mit klaren rechtlichen Nutzungsbedingungen, die sich nach den Parametern von Open Science richten. Besonders wünschenswert wäre es, wenn in diesem Kontext gleich Angaben über verschiedene Stufen der Wieder- und Weiterverwendung inklusive Qualitätslevel der digitalen Daten mitgeliefert werden würden[35], z. B.:

- ***Digitalisat***: Mindestanforderungen für verschiedene Nutzungsszenarien. In einigen Fällen braucht es hochauflösende Digitalisate, in anderen Anwendungsfällen mögen auch niedrigere Auflösungen ausreichend sein.

[30] vgl. DHd AG Datenzentren: Geisteswissenschaftliche Datenzentren im deutschsprachigen Raum - Grundsatzpapier zur Sicherung der langfristigen Verfügbarkeit von Forschungsdaten (Version 1.0). DHd AG Datenzentren, 2018.
[31] DTA - Deutsches Textarchiv: DTA - Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache. http://www.deutschestextarchiv.de/doku/referenzkorpus .
[32] TextGrid: Die Digitale Bibliothek bei TextGrid. https://textgrid.de/digitale-bibliothek.
[33] HathiTrust: Our Digital Library. https://www.hathitrust.org/digital_library.
[34] HathiTrust: Our Research Center. https://www.hathitrust.org/htrc .
[35] vgl. Klaffki / Schmunk / Stäcker 2018. Dort werden verschiedene Digitalisierungsklassen danach unterschieden, inwieweit sie die Nutzung durch digitale Forschungsvorhaben unterstützen.

- ***Volltext***: Mindestanforderungen für verschiedene Nutzungsszenarien. Die Erstellung von Volltexten kann in vielen Fällen (automatisiert) mittels OCR-Verfahren erfolgen, in einigen Anwendungsgebieten braucht es aufwändigere Double-Keying-Methoden. Auch hier ist das Ausgangsmaterial entscheidend.
- ***Metadaten***: Analog zu dem Europeana Data Model oder dem HathiTrust Metadata Management System Zephir braucht es ein Mindest-Kernset an Metadatenelementen, auf die sich die Datenproduzenten gemeinsam mit allen geisteswissenschaftlichen Fachdisziplinen einigen. Dieses Kernset sollte dabei weitgehend bereits etablierte Metadatenschema wie DataCite, Dublin Core etc. nachnutzen, um die Interoperabilität und damit die Nachnutzbarkeit für interdisziplinäre Fragestellungen maximal zu unterstützen.
- ***Kontrollierte Vokabularsysteme***: Für die intellektuelle und automatisierte, maschinenlesbare Nachnutzung braucht es die breite Verwendung von Normdaten, Thesauri, Ontologien etc. über Disziplingrenzen hinweg. Dafür können vorhandene Vokabularsysteme genutzt und/oder erweitert werden, ein Mix verschiedener Systeme oder in einem koordinierten „Kraftakt" ein gemeinsames kontrolliertes Vokabularsystem entwickelt werden, welches auf intelligenten Mappingverfahren zwischen bereits vorhandenen Vokabularsystemen basieren sollte.
- ***Anreicherungen, Verweise, Kommentare, Versionierungen*** etc.: Hierzu zählen Absprachen bezüglich generischer Mindestanforderungen bis hin zu fachwissenschaftlich ausgeprägten Mindeststandards für ein verzweigtes System der tieferen Anreicherung bzw. fachwissenschaftlichen Kommentierung.

Um das Beispiel der Klimaforschung (s. o.) oder der Biodiversität[36] noch einmal zu bemühen, die ebenfalls aus zahlreichen, quer über die großen Wissenschaftsgebiete wie Natur-, Geistes- und Sozialwissenschaften liegenden Teildisziplinen bestehen, besitzen auch die Geisteswissenschaften mit ihren zahlreichen Teildisziplinen und zum Teil sehr unterschiedlichen Praktiken und Kulturen des wissenschaftlichen Arbeitens das große Potential, geeignete, koordinierte Strukturen und Prozesse für eine gemeinsame Forschungsdateninfrastruktur aufzubauen. Diese würde als digitaler Wissensspeicher, als Bewahrinstanz des kulturellen Erbes, gegenüber der fragmentierten und fragilen Landschaft der derzeitigen digitalen Silos einen deutlichen Mehrwert für die Wissenschaft und die Gesellschaft bieten.

Dazu braucht es weniger technisch herausfordernde (Weiter-)Entwicklungen noch riesige Mengen an Speicherkapazität oder Rechnerleistung als vielmehr ein Paradigmenwechsel in der wissenschaftlichen Zusammenarbeit: Eine fachübergreifende Verständigung über Standards quer über alle Aspekte wie Semantik, Syntax, Technologien, Schnittstellen etc. hinweg. So könnte ein umfangreicher, gemeinsamer digitaler Wissensspeicher aufgebaut werden, der für die Bearbeitung von verschiedenen Fragestellungen aus unterschiedlichen (Teil)Disziplinen geeignet ist: Eine neue Kultur des wissenschaftlichen Austauschs, des neugierig Miteinander-Forschen-Wollens entsteht. Es darf bezweifelt werden, ob die Teilchenphysik oder Astrophysik ohne die Notwendigkeit, sich um Großgeräte „gruppieren" zu müssen, heute miteinander weltweit so vernetzt wären und sich auf Standards, Good Practices etc. hätten einigen können. Allein durch die Notwendigkeit, den immensen Kosten dieser Großgeräte einen maximalen Nutzen für eine sehr große Anzahl an beteiligten Wissenschaftler*innen gegenüberzustellen, ist es gelungen, den Gedanken des Teilens von Forschungsdaten innerhalb weniger Jahrzehnte

[36] vgl. German Federation for Biological Data (gfbio). https://www.gfbio.org/.

weltweit so klar zu verbreitern. Dadurch sind auch erst weitere Vorteile der Zugänglichmachung von Forschungsdaten stärker sichtbar geworden wie z. B. Synergieeffekte durch die Vermeidung von Doppelfinanzierungen, Validierung von Forschungsergebnissen, Formulierung ganz neuer Forschungsfragen etc.

Dementsprechend formuliert gleich das zweite Kriterium, welches über die Auswahl der NFDI-Konsortien entscheidet, die Notwendigkeit, einen erwartbaren „Mehrwert für die Entwicklung disziplinübergreifender Metadatenstandards und die Schaffung eines verlässlichen und nachhaltigen Dienste-Angebots im Konsortium“[37] zu schaffen. Es steht also die Frage im Vordergrund, was jede(r) Einzelne dazu beitragen kann, eine geisteswissenschaftliche Forschungsdateninfrastruktur aufzubauen.

3.2 Disziplinübergreifende Handlungsfelder

Die Handlungsfelder für den Aufbau und Betrieb einer geisteswissenschaftlichen Forschungsdateninfrastruktur ergeben sich, neben der fachlich ausgewiesenen Expertise und der nationalen und internationalen Verankerung in den Fach-Communities, insbesondere in der Entwicklung eines stimmigen Konzepts „zu Datennutzung und -zugang sowie Auffindbarkeit und Nachnutzbarkeit der Daten, welches entlang der FAIR-Prinzipien ausgerichtet ist“[38]. Einen guten Überblick über die FAIR-Prinzipien bietet die GO FAIR Initiative[39]. International wird bereits über erste Ansätze zu FAIR-Metriken und -Evaluierungen diskutiert[40], mit deren Hilfe der Grad an „FAIRification“ bewertet werden kann. Zweifellos werden die FAIR-Prinzipien eine immer wichtigere Rolle spielen. Es sei an dieser Stelle betont, dass „FAIR“ nicht mit „Open“ gleichzusetzen ist.

Es lohnt sich, sich die FAIR-Prinzipien im Einzelnen anzuschauen und zu überlegen, was diese für den Diskurs rund um den NFDI-Prozess in den Geisteswissenschaften bedeuten könnten. Die folgende Beschreibung und Liste ist der GO FAIR-Webseite entnommen und führt jeweils zuerst in jedes der vier Haupt-Prinzipien ein, bevor einzelne Unterkriterien genannt werden. „Reusable“ stellt das höchste zu erreichende Ziel dar, da erst hier automatisiert „a la Google“ nach für die Fragestellung geeigneten Daten gesucht werden kann, eine relevante Trefferliste zu erwarten ist und diese Daten sofort unter Beachtung lizenzrechtlicher Bedingungen nachgenutzt werden können:

Findable: The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the FAIRification process.

- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata (defined by R1 below)
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource

[37] GWK 2018, S. 2, Paragraph 5, Absatz b.
[38] GWK 2018, S. 2, Paragraph 5, Absatz g.
[39] vgl. GO FAIR: FAIR Principles. https://www.go-fair.org/fair-principles/.
[40] vgl. zum Beispiel die verschiedenen Arbeits- und Interessensgruppen der internationalen RDA-Initiative unter https://www.rd-alliance.org/groups wie die *FAIR Data Maturity Model WG* oder *BoF - Measuring FAIRness of Digital Objects - RDA 13th Plenary meeting 2019*.

Accessible: Once the user finds the required data, she/he needs to know how can they be accessed, possibly including authentication and authorisation.

- A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
    - A1.1 The protocol is open, free, and universally implementable
    - A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
- A2. Metadata are accessible, even when the data are no longer available

Interoperable: The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data

Reusable: The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

- R1. Meta(data) are richly described with a plurality of accurate and relevant attributes
    - R1.1. (Meta)data are released with a clear and accessible data usage license
    - R1.2. (Meta)data are associated with detailed provenance
    - R1.3. (Meta)data meet domain-relevant community standards[41]

Mit den FAIR-Prinzipien liegen die Handlungsfelder auf der Hand und muss das Forschungsdatenmanagement für die Geisteswissenschaften neu formuliert und etabliert werden. Dabei kommt Metadaten und weiteren standardisierten Beschreibungsformaten (wieder) eine große Bedeutung zu, die sich von „Accessible“ bis zu „Reuse“ durchzieht. Ersichtlich wird auch, dass zur maximalen Unterstützung der FAIR-Prinzipien Absprachen und koordinierte *Good Practices* für eine einheitliche Beschreibung von Forschungsdaten ausschlaggebend sein werden, nicht technische Entwicklungen.

Für die geisteswissenschaftlichen Teildisziplinen und für die Geisteswissenschaften insgesamt müssen daher im Bereich des Forschungsdatenmanagements neue Kulturen und Praktiken gefunden und flächendeckend etabliert werden, die nachfolgende Fragestellungen berücksichtigen:

Im Bereich des ***Datenmanagements*** und der ***Datenmanagementpläne*** (DMPs):

- Welche Tools und Dienste unterstützen das Datenmanagement und die dynamische bzw. aktive Erstellung von DMPs (z. B. *RDMO*)?
- Wie sehen die Anforderungen der jeweiligen Forschungsförderer, Fachgesellschaften, Zeitschriften, Fachdisziplinen etc. aus?
- Was sind gemeinsame Mindestanforderungen zur einheitlichen Beschreibung von Daten, Methoden, Verfahren, wissenschaftlicher Software etc.?
- Welche Mindestqualitätskriterien lassen sich für die einzelnen FAIR-Kriterien definieren? Wie könnten Entwicklungspfade Richtung „Reuse“ aussehen?

[41] GO FAIR: FAIR Principles. https://www.go-fair.org/fair-principles/.

- Welche Fachdisziplin hat in welchem Bereich bereits Vorbildcharakter?

Im Bereich der ***Archivierung*** und ***Publikation***:

- Wie sieht eine gute datenwissenschaftliche Praxis aus? Welche allgemeinen und fachspezifischen Leitfäden lassen sich entwickeln?
- Was muss bzw. kann wie und ab wann zugänglich sein (z. B. *dark archive*, rein technische Langzeitarchivierung, Datenpublikation, Nachweis über Datensammlungen etc.)?
- Welche (zertifizierten, vertrauenswürdigen) Datenzentren und -repositorien gibt es bereits, welche Angebote lassen sich ausbauen und wo bestehen welche (weiteren) Bedarfe?

Im Bereich ***Kompetenzen*** und ***Unterstützung***:

- Wie können (Nachwuchs)Wissenschaftler*innen im Bereich gute datenwissenschaftliche Praxis unterstützt und geschult werden?
- Wer kann sich wie um das Forschungsdatenmanagement kümmern? Welche neuen Karrierepfade und Berufsfelder entstehen in diesem Kontext? Wie kann dies durch geeignete Job-Beschreibungen und Aus- sowie Weiterbildungsangebote unterstützt werden?

Übergreifende Fragestellungen, die zumindest eine einheitliche Vorgehensweise verlangen:

- Sollen pro Disziplin Strategien, Policies, Leitlinien etc. entwickelt werden bzw. kann es z. B. als Referenzrahmen eine übergreifende geisteswissenschaftliche Policy geben, die auch Aspekte des digitalen kulturellen Erbes berücksichtigt?
- Sollten Strategien, Policies etc. eher vom Datentyp her entwickelt, also eine Ausdifferenzierung in Text, Bild, Ton und Video vorgenommen werden?
- Was ist der State-of-the-Art hinsichtlich Werkzeuge, digitale Dienste, Datenzentren, Zitierpraktiken, Datenpublikationen, Umgang mit wissenschaftlicher Software etc.?
- Wie könnte ein Referenzcurriculum für das Forschungsdatenmanagement und für den Umgang mit digitalen Daten im Allgemein aussehen? Welche Bausteine müssten grundlegend in die geisteswissenschaftliche Lehre (Bachelor und Master) integriert werden, um dem Anspruch „Investition in Köpfe“[42] gerecht zu werden? Welche Grundkenntnisse in z. B. Statistik, Algorithmen, Technologien, Programmierung etc. sind für zukünftige Geisteswissenschaftler*innen unerlässlich?
- Wie können zukünftige Anreizsysteme, Anerkennungsverfahren und Karrierepfade entwickelt werden, um einer neuen Generation von Geisteswissenschaftler*innen das Arbeiten mit digitalen Daten inklusive der Publikation, der Entwicklung wissenschaftlicher Software, der Unterstützung des Forschungsdatenmanagements etc. als eigenständige wissenschaftliche Leistung zu vermitteln? Was kann auch hier von anderen Fachdisziplinen gelernt werden, die zum Teil Datenpublikationen in renommierten Fachzeitschriften mit über 100 Autor*innen veröffentlichen oder bei denen sogenannte *data manager* oder *data steward*s bereits etablierte Berufszweige sind?

[42] RfII - Rat für Informationsinfrastrukturen: Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland. Göttingen 2016, S. 36.

4. Fazit & Ausblick

Insgesamt stehen die Geisteswissenschaften vor großen Herausforderungen. Im Rahmen des europäischen Prozesses um die *European Open Science Cloud* (EOSC) und des von Bund und Ländern gemeinsam initiierten NFDI-Programms bietet sich jedoch eine große Chance, sowohl die Geisteswissenschaften im digitalen Zeitalter neu zu positionieren als auch auf die Bedeutung digitaler Daten(sammlungen) auf eine andere Art und Weise aufmerksam zu machen.

Durch die massive Förderung seitens der Europäischen Kommission und deutscher Förderorganisationen (z. B. DFG, BMBF, Stifterverband) wird die Bedeutung von digitalen Forschungsdaten und ihrem sachgemäßen (FAIRen) Umgang schnell und rasant wachsen. Auch der kommerzielle Dienstleister Google hat sich bereits der Thematik angenommen und eine Beta-Version seiner Dataset Search veröffentlicht, die es Nutzern ermöglichen soll, digitale Datensätze bzw. Datensammlungen mittels einer einfachen Stichwortsuche zu finden. Inwieweit lassen sich die Ziele von Google mit wissenschaftlichen Anforderungen wie den FAIR-Prinzipien vereinbaren? Müssen auch solche Initiativen in Bestrebungen für eine bessere Findbarkeit von Daten einbezogen werden?[43]

Während viele Fachdisziplinen außerhalb der Geisteswissenschaften in den letzten Jahrzehnten umfangreiche, wertvolle Datensammlungen aufgebaut haben (z. B. in den Sozialwissenschaften das *Sozio-ökonomisches Panel*) bzw. gerade aufbauen (z. B. in der Teilchenphysik durch das FASER-Experiment[44] am CERN), stehen ähnliche Bestrebungen für den Aufbau umfangreicher geisteswissenschaftlicher Datensammlungen in Deutschland noch am Anfang. Hierdurch bietet sich die Gelegenheit des Aufbaus eines großen, digitalen Wissensspeichers, der einheitlich durchsucht und bei dem ein eigenes Forschungskorpus heruntergeladen und nachgenutzt werden kann. Dieser „Referenzkorpus“ könnte dazu beitragen, die Bedeutung der Geisteswissenschaften wieder fest in der Gesellschaft, bei den Bürger*innen, zu verankern.

Zur Realisierung des Korpus sind sicherlich mehrere iterative Schritte nötig. Einerseits sollte es von Anfang an ein niedrigschwelliges Angebot für alle Nutzer*innen einschließlich interessierten Laien sein, in den Datensammlungen zu stöbern, um z. B. literaturwissenschaftliche (oder völlig neue) Fragestellungen digital zu erkunden, anderseits könnten gezielt disziplinär interessierte Fachwissenschaftler*innen zu Kooperation und gemeinsamen, schrittweisen Entwicklung neuer Forschungsfragen und Nutzungsszenarien motiviert werden.

In den letzten Jahrzehnten ist es einigen Fachdisziplinen gelungen, zum Teil spektakuläre Fördersummen für ihre Forschungsanliegen und die Entwicklung und Bereitstellung der dafür notwenigen Forschungsinfrastrukturen einzuwerben. So gelang es der Klimaforschung als eine der jüngsten Fachdisziplinen, die auch immer wieder oder gerade wieder sehr heftig in die Kritik gerät und deren Forschungsergebnisse und -daten abgestritten oder ignoriert werden, sich durch eine gemeinsame Koordinierung und internationale Anschlussfähigkeit nicht nur innerhalb kürzester Zeit auf vorbildliche Art und Weise quer über verschiedene Fachdisziplinen[45] auf internationaler Ebene zu vernetzen, sondern auch durch die Interessengemeinschaft *Deutsches Klima Konsortium* (DKK) mit einer Stimme gegenüber Politik, Gesellschaft, Förderer

[43] vgl. Blumtritt, Jonathan / Rau, Felix: Metadaten im Zeitalter von Google Dataset Search. Zenodo 2019.
[44] vgl. z. B. Gast, Robert: CERN bewilligt neuen LHC-Detektor. Spektrum.de News 06.03.2019. https://www.spektrum.de/news/cern-bewilligt-neuen-lhc-detektor/1627998.
[45] Zu nennen sind hier neben Naturwissenschaften wie Meteorologie, Ozeanographie, Chemie und Physik auch Sozialwissenschaften wie Politologie, Ökonomie, Soziologie etc. sowie Technik- und Ingenieurswissenschaften.

etc. aufzutreten. Auch die rapide Entwicklung geeigneter (fachwissenschaftlicher) Standards, Leitfäden, Policies und der vorbildliche Umgang mit den eigenen Forschungsdaten zeigt in diesem Fall, wie eine Fachdisziplin eine Vorreiterrolle einnehmen kann. Die Geisteswissenschaften bilden diesbezüglich eher das Schlusslicht, nicht nur was die Einwerbung von Fördermitteln für digitale geisteswissenschaftliche Infrastrukturen betrifft. Die Investitionshöhe ist im Vergleich zu anderen Wissenschaften vergleichsweise gering. Dies liegt sicherlich auch daran, dass sie sich nicht so durchsetzen können, weil sie viel eher geneigt und gewohnt sind, die Unterschiede zwischen ihren Fächerkulturen zu betonen als mit „einer gemeinsamen Stimme zu sprechen".

Ohne abzustreiten, dass die Herausforderungen in den geisteswissenschaftlichen Disziplinen aufgrund disziplinärer Unterschiede vielfältig sind und eine lange fragmentierte Praxis in den (Teil)Disziplinen die Konsensbildung verzögert bzw. behindert, wäre der erste wichtige Schritt eine detaillierte Kartierung der geisteswissenschaftlichen Fächer, um den Status Quo in den Bereichen geeigneter Vorbilder und (gemeinsamer) Bedarfe zu identifizieren, für die dann in vernetzter und abgestimmter Weise weitere gemeinsame Lösungen gefunden werden. Dabei sollte auch die Frage nach der Rolle der Gedächtniseinrichtungen (Bibliotheken, Museen, Archive etc.) als Informationsinfrastrukturanbieter im Zusammenspiel mit den Geisteswissenschaften einbezogen werden. Bei Fragen von Betriebsmodellen, Nachhaltigkeit, Vertrauenswürdigkeit etc. besitzen die geisteswissenschaftlichen Fächer bereits eine seit Jahrhunderten funktionierende Infrastruktur durch die Gedächnisinstitutionen, die in den anderen, größtenteils natur- und sozialwissenschaftlichen Disziplinen erst teuer und mühsam aufgebaut werden muss(te). So stehen Bibliotheken, Museen und Archive als Sinnbild für neutrale, vertrauenswürdige und langlebige Einrichtungen, deren physische und mehr und mehr auch digitale Infrastruktur den Forschenden und Studierenden langfristig zur Verfügung steht.

Insgesamt kann die eingangs gestellte Frage nach der Utopie oder Dystopie einer gemeinsamen geisteswissenschaftlichen Forschungsdatenkultur zum jetzigen Zeitpunkt nicht abschließend beantwortet werden. Es bleibt abzuwarten, wie sich die geisteswissenschaftlichen Fächer im NFDI-Prozess positionieren werden und inwieweit es gelingen wird, über Traditionsgrenzen hinweg neue Formen der Kooperation aufzubauen und nachhaltig mit Leben zu füllen. Die Fachgesellschaften können dabei eine hilfreiche und unterstützende Rolle einnehmen, allerdings haben sie sich in der Vergangenheit beim Themenkomplex Forschungsdatenmanagement nicht besonders engagiert gezeigt. Vielleicht gelingt es den neueren Fachverbänden bzw. Interessensgemeinschaften wie dem *Verband Digital Humanities im deutschsprachigen Raum* e.V. (DHd) oder dem neuen *Geisteswissenschaftlichen Forum NFDI* eine neue Kultur der Kooperation zu initiieren.

Nicht umsonst heißt es in der Präambel der *Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI)*: „Dazu muss das Datenmanagement standardisiert sein: Nach den sogenannten FAIR-Prinzipien sollen Forschungsdaten auffindbar, zugänglich, interoperabel und nachnutzbar sei". Es sollen also NFDI-Konsortien gefördert werden, die glaubhaft machen können, dass sie sich auf den Weg zu FAIRen Daten begeben und darlegen können, wie sie sich die Umsetzung in den nächsten zehn Jahren vorstellen. Dies ist eine gewaltige Aufgabe für die Geisteswissenschaften und ihre tradierten Verfahren, heterogenen Fächerkulturen und fragmentierten Disziplinen, aber wiederum eine enorme Chance für ein großes, geisteswissenschaftliches „Daten-Labor" als Äquivalent zum Großgerät (z. B. in der Astrophysik).

Falls nur einige der nachfolgenden Fragen mit „Ja“ beantwortet werden können, könnten sich die Geisteswissenschaften glaubhaft und zukunftsorientiert eine eigene Forschungsdateninfrastruktur aufbauen: Besteht ein gemeinsamer Nenner, um geisteswissenschaftliche Sammlungen zu „atomisieren“ (Korpora, Beschreibungen, Metadaten) damit diese als „Großgeräte“ und „Labore“ und Ausgangspunkt für verschiedene Fragestellungen dienen können? Macht es Sinn, für jede einzelne geisteswissenschaftliche Disziplin ein eigenes Daten-Repository aufzubauen?[46] Oder braucht es nur wenige, international vernetzte, Repositorien in Deutschland (vgl. HathiTrust Digital Library)? Und natürlich muss auch geklärt werden, wo und ab wann disziplinäre Grenzen eine Rolle spielen oder ob sich die Daten einer objektiven Atomisierung verweigern, da sie auf subjektiven Interpretationen beruhen.

Literaturverzeichnis

Blumtritt, Jonathan / Rau, Felix: Metadaten im Zeitalter von Google Dataset Search. Zenodo 2019. DOI: http://doi.org/10.5281/zenodo.2613444 (Zugriffsdatum: 24.04.2019).

BMBF - Bundesministerium für Bildung und Forschung: Forschung und Innovation für die Menschen: Die Hightech-Strategie 2025. Berlin 2018. https://www.bmbf.de/upload_filestore/pub/Forschung_und_Innovation_fuer_die_Menschen.pdf (Zugriffsdatum: 24.04.2019).

Borgman, Christine L.: Big Data, Little Data, No Data. Scholarship in the Networked World. Cambridge (Mass.), London: MIT Press 2015.

Buddenbohm, Stefan / Engelhardt, Claudia / Wuttke, Ulrike: Angebotsgenese für ein geisteswissenschaftliches Forschungsdatenzentrum. In: Zeitschrift für digitale Geisteswissenschaften 2016,1. DOI: http://dx.doi.org/10.17175/2016_003 (Zugriffsdatum: 24.04.2019).

Cremer, Fabian / Klaffki, Lisa / Steyer, Timo: Der Chimäre auf der Spur: Forschungsdaten in den Geisteswissenschaften. In: O-Bib 5,2 (2018), S. 142-162. DOI: https://doi.org/10.5282/o-bib/2018H2S142-162 (Zugriffsdatum 24.04.2019).

DARIAH-DE. https://de.dariah.eu/ (Zugriffsdatum: 24.04.2019).

DARIAH-DE: Weiterführende Informationen. https://de.dariah.eu/weiterfuhrende-informationen (Zugriffsdatum: 24.04.2019).

DataCite Metadata Schema. https://schema.datacite.org/ (Zugriffsdatum: 24.04.2019).

Deutsches Klimarechenzentrum (DKRZ). https://www.dkrz.de/ (Zugriffsdatum: 24.04.2019).

DFG - Deutsche Forschungsgemeinschaft: Vorschläge zur Sicherung guter wissenschaftlicher Praxis / Proposals for Safeguarding Good Scientific Pratice, Denkschrift / Memorandum, Empfehlungen der Kommission „Selbstkontrolle in der Wissenschaft“ / Recommendations of the Commission on Professional Self Regulation in Science. Bonn: Wiley-VCH 2013. https://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf (Zugriffsdatum: 24.04.2019).

[46] vgl. z. B. DHd-AG Datenzentren 2017.

DFG – Deutsche Forschungsgemeinschaft: Leitlinien zum Umgang mit Forschungsdaten. 2015. https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/richtlinien_forschungsdaten.pdf (Zugriffsdatum: 24.04.2019).

DFG - Deutsche Forschungsgemeinschaft: Umgang mit Forschungsdaten. https://www.dfg.de/foerderung/antrag_gutachter_gremien/antragstellende/nachnutzung_forschungsdaten/ (Zugriffsdatum: 24.04.2019).

DHd - Digital Humanities im deutschsprachigen Raum. https://dig-hum.de/ (Zugriffsdatum: 24.04.2019).

DHd AG Datenzentren: Geisteswissenschaftliche Datenzentren im deutschsprachigen Raum - Grundsatzpapier zur Sicherung der langfristigen Verfügbarkeit von Forschungsdaten (Version 1.0). DHd AG Datenzentren, 2018. DOI: http://doi.org/10.5281/zenodo.1134760 (Zugriffsdatum: 24.04.2019).

DKK - Deutsches Klima-Konsortium. https://www.deutsches-klima-konsortium.de/ (Zugriffsdatum: 24.04.2019).

DTA - Deutsches Textarchiv. http://www.deutschestextarchiv.de (Zugriffsdatum: 24.04.2019).

DTA - Deutsches Textarchiv. DTA - Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache. http://www.deutschestextarchiv.de/doku/referenzkorpus (Zugriffsdatum: 24.04.2019).

Dublin Core Metadata Element Set, Version 1.1: Reference Description. http://www.dublincore.org/specifications/dublin-core/dces/ (Zugriffsdatum: 24.04.2019).

EOSC - European Open Science Cloud. https://www.eosc-portal.eu/ (Zugriffsdatum; 24.04.2019).

European Commission Expert Group on FAIR Data: Turning FAIR into Reality. Final Report and Action Plan from the European Commission Expert Group on FAIR Data. European Commission: Brüssel 2018. DOI: https://doi.org/10.2777/1524 (Zugriffsdatum: 24.04.2019).

Europeana: Europeana Data Model. https://pro.europeana.eu/resources/standardization-tools/edm-documentation (Zugriffsdatum: 24.04.2019).

FORCE11: The FAIR Data Principles. https://www.force11.org/group/fairgroup/fairprinciples (Zugriffsdatum: 24.04.2019).

Gast, Robert: CERN bewilligt neuen LHC-Detektor. Spektrum.de News 06.03.2019. https://www.spektrum.de/news/cern-bewilligt-neuen-lhc-detektor/1627998 (Zugriffsdatum: 24.04.2019).

Geisteswissenschaftliches Forum NFDI. https://nfdi.hypotheses.org/ (Zugriffsdatum: 24.04.2019).

German Federation for Biological Data (gfbio). https://www.gfbio.org/ (Zugriffsdatum 24.04.2019).

GO FAIR: FAIR Principles. https://www.go-fair.org/fair-principles/ (Zugriffsdatum: 24.04.2019).

Google: Google Data Set Search Beta, Webseite. https://toolbox.google.com/datasetsearch (Zugriffsdatum: 24.04.2019).

GWK - Gemeinsame Wissenschaftskonferenz: Bund-Länder-Vereinbarung zu Aufbau und Förderung einer Nationalen Forschungsdateninfrastruktur (NFDI) vom 26. November 2018 (BAnz AT 21.12.2018 B10). https://www.gwk-bonn.de/fileadmin/Redaktion/Dokumente/Papers/NFDI.pdf (Zugriffsdatum: 24.04.2019).

GWK - Gemeinsame Wissenschaftskonferenz: Informationsinfrastrukturen / NFDI. https://www.gwk-bonn.de/themen/weitere-arbeitsgebiete/informationsinfrastrukturen-nfdi/ (Zugriffsdatum: 24.04.2019).

HathiTrust: Our Digital Library. https://www.hathitrust.org/digital_library (Zugriffsdatum: 24.04.2019).

HathiTrust: Our Research Center. https://www.hathitrust.org/htrc (Zugriffsdatum: 24.04.2019).

HathiTrust: Zephir. https://www.hathitrust.org/zephir (Zugriffsdatum: 24.04.2019).

Klaffki, Lisa / Schmunk, Stefan / Stäcker, Thomas: Stand der Kulturgutdigitalisierung in Deutschland: Eine Analyse und Handlungsvorschläge des DARIAH-DE Stakeholdergremiums „Wissenschaftliche Sammlungen“, Göttingen 2018 (DARIAH-DE Working Papers, 26). URN: http://nbn-resolving.de/urn:nbn:de:gbv:7-dariah-2018-1-3 (Zugriffsdatum: 24.04.2019).

Institute of Museum and Library Services (IMLS). https://www.imls.gov/ (Zugriffsdatum: 24.04.2019).

Knoche, Michael: Die Idee der Bibliothek und ihre Zukunft. Göttingen: Wallstein 2018.

Lingenhöhl, Daniel: Der tiefe Fall eines Ernährungspapstes. Spektrum.de News 06.03.2019. Webseite. https://www.spektrum.de/news/der-tiefe-fall-eines-ernaehrungspapstes/1593004 (Zugriffsdatum: 24.04.2019).

Max-Planck-Institut für Astronomie: Startschuss für modernste Kartierung des Himmels. Webseite. http://www.mpia.de/news/wissenschaft/2000-03-sdss (Zugriffsdatum: 24.04.2019).

NASA Science: Astrophysics Data Centers. https://science.nasa.gov/astrophysics/astrophysics-data-centers (Zugriffsdatum: 24.04.2019).

Neuroth, Heike / Engelhardt, Claudia / Klar, Jochen / Enke, Harry: Aktives Forschungsdatenmanagement - Research Data Management Organiser. In: ABI Technik 38,1 (2018), S. 55-64. DOI: http://doi.org/10.1515/abitech-2018-0008 (Zugriffsdatum: 24.04.2019).

NOAA - National Oceanic and Atmospheric Administration - Earth System Research Laboratory (ESRL) - Physical Sciences Division (PSD): Data and Imagery. https://www.esrl.noaa.gov/psd/data/ (Zugriffsdatum: 24.04.2019).

Open Research Data (ORD). https://data.europa.eu/euodp/de/data/dataset/open-research-data-the-uptake-of-the-pilot-in-the-first-calls-of-horizon-2020 (Zugriffsdatum: 24.04.2019).

Open Research Data Task Force: Realising the Potential. Final Report of the Open Research Data Task Force“. Open Research Data Task Force, 2018. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/775006/Realising-the-potential-ORDTF-July-2018.pdf (Zugriffsdatum: 24.04.2019).

Pronk, Tessa E.: The Time Efficiency Gain in Sharing and Reuse of Research Data, In: Data Science Journal 18,1 (2019), S. 10. DOI: http://doi.org/10.5334/dsj-2019-010 (Zugriffsdatum: 24.04.2019).

Raspe, Martin / Schelbert, Georg: Genau, wahrscheinlich, eher nicht. Beziehungsprobleme in einem kunsthistorischen Wissensgraph. In: Die Modellierung des Zweifels – Schlüsselideen und -konzepte zur graphbasierten Modellierung von Unsicherheiten. Hg. von Andreas Kuczera / Thorsten Wübbena / Thomas Kollatz. 2019 (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 4). DOI: http://dx.doi.org/10.17175/sb004_012 (Zugriffsdatum: 24.04.2019).

RDA DE: RDA Deutschland Tagung 2019. https://www.rda-deutschland.de/events/tagung-2019 (Zugriffsdatum: 24.04.2019).

RDMO - Research Data Management Organiser. https://rdmorganiser.github.io/ (Zugriffsdatum: 24.04.2019).

RfII - Rat für Informationsinfrastrukturen: Leistung aus Vielfalt. Empfehlungen zu Strukturen, Prozessen und Finanzierung des Forschungsdatenmanagements in Deutschland. Göttingen 2016. http://www.rfii.de/?p=1998 (Zugriffsdatum: 24.04.2019).

Sahle, Patrick / Kronenwett, Simone: Jenseits der Daten. Überlegungen zu Datenzentren für die Geisteswissenschaften am Beispiel des Kölner ‚Data Center for the Humanities’. In: LIBREAS. Library Ideas, 23 (2013), S. 76–96. http://dx.doi.org/10.18452/9043 (Zugriffsdatum: 24.04.2019).

Schöch, Christof: Aufbau von Datensammlungen. In: Jannidis, Fotis / Hubertus Kohle / Malte Rehbein (Hg). Digital Humanities. Eine Einführung. Stuttgart: J.B. Metzler, 2017, S. 223-233.

Schwerpunktinitiative „Digitale Informationen“: Forschungsdaten. https://www.allianzinitiative.de/de/archiv/forschungsdaten/ (Zugriffsdatum: 24.04.2019).

SDSS - Sloan Digital Sky Survey. https://www.sdss.org/ (Zugriffsdatum: 24.04.2019).

SDSS - Sloan Digital Sky Survey. How to Cite SDSS. https://www.sdss.org/collaboration/citing-sdss/ (Zugriffsdatum: 24.04.2019).

SOEP - Sozio-oekonomische Panel. https://www.diw.de/soep (Zugriffsdatum: 24.04.2019).

TextGrid: Die Digitale Bibliothek bei TextGrid. Webseite. https://textgrid.de/digitale-bibliothek (Zugriffsdatum: 24.04.2019).

The Santa Barbara Statement on Collections as Data. https://collectionsas-data.github.io/statement/ (Zugriffsdatum: 24.04.2019).

ZVDD - Zentrales Verzeichnis Digitalisierter Drucke. http://www.zvdd.de/startseite/ (Zugriffsdatum: 24.04.2019).

Wilkinson, Mark D. et al.: The FAIR Guiding Principles for Scientific Data Management and Stewardship. in: Scientific Data 3 (2016). https://doi.org/10.1038/sdata.2016.18 (Zugriffsdatum: 24.04.2019).

Ist da eine Veränderung in den Geisteswissenschaften? Infrastrukturen und ihre Folgen für die Praxis der Geisteswissenschaften

Gerhard Lauer (Basel)

Im zweiten Buch seiner Abhandlung *The Advancement of Learning* von 1605 hat Francis Bacon die Infrastrukturen für die zu gründenden modernen Wissenschaften umrissen. Für das, was wir heute näherungsweise Geisteswissenschaften nennen, nennt er zwei wesentliche Voraussetzungen gelingender Wissenschaften. Bacon schreibt:

> The works touching books are two — first, libraries, which are as the shrines where all the relics of the ancient saints, full of true virtue, and that without delusion or imposture, are preserved and reposed; secondly, new editions of authors, with more correct impressions, more faithful translations, more profitable glosses, more diligent annotations, and the like.[1]

Bacon zählt mit guten Gründen unter die Gründungsväter der modernen Wissenschaften im konfessionellen Zeitalter.[2] Infrastrukturen für die neuen Wissenschaften, das waren für Bacon Bibliotheken und Editionen. Sie ermöglichen, was wir heute zumeist Geisteswissenschaften nennen und das umso mehr, wenn man unter Bibliotheken auch Archive und Museen, unter Editionen auch Sammlungen und die vielen Formen des Weiterschreibens und Übersetzens in Büchern, Zeitschriften bis zum Feuilleton versteht, die geisteswissenschaftliches Arbeiten vielfach ausmachen. Man sieht unschwer, dass die geisteswissenschaftlichen Infrastrukturen zunächst buchzentriert waren. Ohne Bibliotheken und Editionen sind Geisteswissenschaften kaum eine Praxis und keine Disziplinen und das bis heute. Dass den buchzentrierten Fächern der Theologie und der Klassischen Philologie lange eine historische Vorrangstellung zukam, ist daher kein historischer Zufall. Sie sind die Buchwissenschaften schlechthin.
Doch Infrastrukturen waren in der Frühen Neuzeit bald schon mehr als Bibliotheken und Editionen. Hinzu kamen Bibliographien und Rezensionsorganen wie die *Göttingischen Gelehrten Anzeigen*, Akademien für die Forschung, da in den Universitäten nicht geforscht, sondern gelehrt wurde. Dazu kamen auch gelehrte Korrespondenznetzwerke und schließlich auch gegen Ende des 19. Jahrhunderts größere, lang angelegte Forschungsprojekte wie Boeckhs *Corpus Inscriptionum Graecarum*, Mommsens epigraphische und prosopographische Projekte, die *Monumenta Germaniae Historica*, Kirchenväterkommission, *Prosopographie Imperii Romani* oder das *Corpus Inscriptionum Latinarum*.[3] Diese Projekte wurden in enger Nachbarschaft und zugleich Konkurrenz zu der entstehenden Großforschung in den Naturwissenschaften konzipiert. Mommsen und Harnack, Geistes- und Naturwissenschaften teilten die identischen forschungspolitischen Strukturen des Kaiserreichs.

[1] vgl. Bacon, Francis: The Advancement of Learning. In: Francis Bacon. The Major Works. Edited with an Introduction and Notes by Brian Vickers. Oxford: Oxford University Press 1996 [1605], S. 170.
[2] vgl. Gaukroger, Stephen: Francis Bacon and the Transformation of Early-Modern Philosophy. Cambridge: Cambridge University Press 2010.
[3] vgl. Rebenich, Stefan: Theodor Mommsen und Adolf Harnack. Wissenschaft und Politik des ausgehenden 19. Jahrhunderts. Berlin: de Gruyter 2012.

Forschungsinfrastrukturen waren daher schon zu Beginn des 20. Jahrhunderts mehr als nur Editionen und Bibliotheken, waren bald schon nicht nur buchzentriert. Sie sind es noch weniger als mit den Massenuniversitäten des 20. Jahrhunderts auch die Geisteswissenschaften weit mehr als nur Editionen und Bibliotheken brauchen, um zu funktionieren. Der Humboldt-Kult der 20er Jahre des 20. Jahrhunderts war schon eine Reaktion darauf, dass auch die Geisteswissenschaften kaum den idealistischen Idealen der Weimarer Klassik entsprachen als sie zwischen den Kriegen ihren Status ausbauen konnten. Das Pathos von Einsamkeit und Freiheit, das damals so gerne beschworen wurde und sich bis heute in den bildungspolitischen Debatten gehalten hat, war eine ideologische Konstruktion vor hundert Jahren und die Humboldt-Edition von Eduard Spranger u.a. ist ein Zeugnis für diese Selbstverklärung der Geisteswissenschaften angesichts von Großforschung und – wie man damals schon es empfand – Massenuniversität der 20er Jahre.[4] Wenn schon im 19. und erst recht im 20. Jahrhundert Infrastrukturen komplexe Institutionen waren, um wieviel weniger passen Sentimentalitäten vom guten Buch und den einfachen Infrastrukturen zu den Geisteswissenschaften des 21. Jahrhunderts.

1. Transformation der Geisteswissenschaften

Damit bin ich bei meiner ersten These: Die Digitalisierung verstanden als Transformation einer Wissenschaft durch digitale Ressourcen und veränderte Methoden, ihrer anderen Darstellung und Vernetzung in der Gesellschaft, diese Digitalisierung ist nur ein, wenn auch gewichtiger Faktor in der Umgestaltung der Wissenschaften. Nicht alles ist neu, was unter Digitalisierung ausgeflaggt wird, aber vieles schon und das hat längerfristige Folgen, von denen hier zu reden sein wird. Dennoch wäre es eine grobe Vereinfachung, Faktoren wie die veränderten Bildungszugänge, die Globalisierung der Wissenschaften, ihre interne Dynamik der Ausdifferenzierung, die massiven ökonomischen Interessen und andere Faktoren außen vor zu lassen oder diese unter Digitalisierung zu subsumieren. Moderne Gesellschaften sind multifaktoriell bestimmt. Wenn daher von der Digitalisierung, hier den digitalen Infrastrukturen, die Rede sein muss, ist darauf zu achten, den Entdifferenzierungen und Simplifikationen aus dem Weg zu gehen, die sich nur zu leicht einstellen, wenn von Digitalisierung die Rede ist. Das Schlagwort ‚Digitalisierung' tendiert zur Simplifikation. Es spricht viel dafür, die damit implizierten ideologisierten Debatten zu vermeiden.

Was also heißt und zu welchem Ende reden wir von digitalen Infrastrukturen? Ich schichte das Problem ab, indem ich versuche differenzierter zu fragen: Welche Veränderungen im Gegenstandsfeld der Geisteswissenschaften gibt es und was zeitigt Folgen für die Infrastrukturen der Geisteswissenschaften? Welche Veränderungen der Methoden ziehen Veränderungen in den Infrastrukturen nach sich? Wie verändern sich Rollen und Akteure im geisteswissenschaftlichen Forschungs- und Lehrbetrieb, wie ändern sich die Umwelten, so dass Infrastrukturen für die Geisteswissenschaften des 21. Jahrhunderts andere sind als die für das 20. Jahrhundert.

[4] vgl. Gerhard (Hg.): Wilhelm von Humboldt. Schriften zur Bildung. Stuttgart: Reclam 2017.

2. Infrastrukturen und die Gegenstände der Geisteswissenschaften

Das Gegenstandsfeld der Geisteswissenschaft unterlag und unterliegt einer Erweiterung längst vor der Digitalisierung. Die Ausdehnung von den klassischen und biblischen Gegenständen auf mittelalterliche Überlieferungen im 19. Jahrhundert und bald schon auch auf außereuropäische Gegenstandsbereiche hat viele neue Fächer wie die Orientalistik, Keltologie oder Germanistik erfunden. Fächer wie die Amerikanistik oder Neugermanistik kamen im 20. Jahrhundert dazu, weil sich die Relevanz von Fragestellungen gewandelt hatte. Die Kulturwissenschaften haben schließlich das Gegenstandsfeld geisteswissenschaftlicher Fächer soweit ausgedehnt, dass in vielen Fächer nur noch wenig Konsens darüber herrscht, was Gegenstand des jeweiligen Fachs ist. Ist in der Anglistik der Fahrplan der Londoner Tube, in der Archäologie auch die Wissenschaftspopularisierung in der Museumskultur der Gegenwart ein disziplinär akzeptierter Gegenstand? Die Antworten auf solche und ähnliche Fragen fallen unterschiedlich aus.

Die Reihe lässt sich leicht fortsetzen, wenn es etwa Debatten darüber gibt, ob *Latino studies* und *Latina studies* zwei verschiedene Gegenstandsfelder und evtl. auch Disziplinen meinen könnten. Dem teilweise bizarren Erfindungsreichtum der Geisteswissenschaften für neue Gegenstandsfelder lassen sich leicht reichlich ironische Wendungen abgewinnen. Dennoch lohnt es sich genauer hinzusehen. Gegenstandsfelder sind disziplinäre Konstruktionen, kein Zweifel, und sie folgen binnenwissenschaftlichen wie außerwissenschaftlichen Faktoren.[5] Digitalisierung ist daher kein notwendiger Gegenstand der Geisteswissenschaften und nicht wenige Kolleginnen und Kollegen sehen das genau so. Dass es mehr Kunst, Musik und Literatur außerhalb des Kanons gibt, führt eher zu einem Unbehagen in der Kultur. Die Reaktion sind dann mehr oder weniger gut begründete Varianten der Kulturkritik vom Niedergang der Bildung, dem moralischen Verfall der Jugend und dem Ende der Kultur. Das kaum noch jemand lesen, ist die gängigste Variante dieser Kulturkritik. Und sie führt zu einer stillen Selbstabschaffung der Geisteswissenschaften, der die Studenten wegbleiben.

Es ist also eine in jedem Fach, aber auch in den Fakultäten und den Universitäten jeweils zu treffende Entscheidung, ob die objektiven Veränderungen in der Kultur disziplinäre Veränderungen in den Geisteswissenschaften und ihren Infrastrukturen nach sich ziehen. Mit objektiven Veränderungen im Gegenstandsfeld der Geisteswissenschaften meine ich zunächst die digitale Massenerschließung des kulturellen Erbes von der Keilschrift-Überlieferung in Portalen wie *CDLI* bis zu Bildersammlungen der Getty-Foundation, der Maya-Inschriften bis zur digitalen Mozart-Ausgabe. Das dehnt das mögliche Gegenstandsfeld der Fächer erheblich aus und vertieft deren Forschungsmöglichkeiten zugleich. Die Breite der erschlossenen Überlieferung wie auch deren Tiefe, genauer die Granularität, mit der Details etwa von Keilschrift-Inschriften erfasst werden können, so dass anhand hochauflösender Photographien Zusammenhänge zwischen Tontafeln sichtbar werden, die für das bloße Auge nicht erkennbar sind, das alles ist eine neue Qualität im Gegenstandsfeld.

Keine dieser Erschließungen und Editionen ist möglich ohne eine mehr als proprietäre Infrastruktur. *FuD* in Trier und das *DaSCH* in Basel sind virtuelle Forschungsumwelten für Editionen,

[5] vgl. Diemer, Alwin: Art. Geisteswissenschaften. In: Historisches Wörterbuch der Philosophie 3 (1974), S. 211-215, sowie Hollinger, David. A.: The Humanities and the Dynamics of Inclusion since World War II. Baltimore: Johns Hopkins University Press 2006.

in die auch ältere Pioniereditionen wie etwa Walter Morgenthalers Gottfried Keller-Ausgabe gehoben werden. Standards wie TEI und jetzt in der Musikwissenschaft MEI haben sich aus dieser Ausdehnung des Gegenstandsfelds ergeben und weltweit geteilte Standards haben sich dank der Beharrlichkeit einer kleinen Gruppe von Kolleginnen und Kollegen entwickelt. Das wird durch höherstufige Infrastrukturen wie *DARIAH* und *CLARIN* unterstützt, die helfen, digitale Ressourcen nach gemeinsamen Standards zu archivieren, zu teilen und Werkzeuge zu deren Erschließung einzubinden. Hat man diesen Pfad einmal eingeschlagen, dann hat das wiederum Rückwirkungen auf das Gegenstandsfeld. Denn statt nur Editionen in der Tradition der Werkausgaben zu betreiben, werden Gegenstandsfelder unter Portalen wie etwa *Perseus* zusammengeführt, welches die gesamte schriftliche Überlieferung der Antike und ihre Rezeptionsgeschichte vereint. Genauer: Aus Werken werden Korpora. In der Linguistik hat dies ein ganzes Fach, eben die Korpuslinguistik, hervorgebracht. In kleinen Fächern wie der Koptologie, die nur selten die Werkästhetik der großen Fächer kennen, führt dies zu einer Neuvermessung ihrer Gegenstände. Statt nur die wenigen, traditionell hochbewerteten Gegenstände in den Blick zu nehmen, werden historisch lange Reihen vielfach nur höchst fragmentarischer Überlieferungen zum Gegenstand solcher Fächer. Der Gegenstandsbereich dieser Fächer verändert sich damit nachhaltig. Gerade für die kleinen Fächer ist daher die Digitalisierung ihrer Gegenstände die Chance für eine fachpolitische Erneuerung in Richtung auf eine Vertiefung ihrer Gegenstände und deren Zusammenhänge. Mehr Gegenstände in einer höheren Auflösung und in einer historischen Vernetzung, die noch vor kurzem unmöglich schienen, beginnt die Fächer und deren Gewichte zu verschieben, vorausgesetzt, ihnen gelingt der Aufbau von fachadäquaten Infrastrukturen.

Das ist im Detail dann aufwändig und nicht alle Fächer verfügen dafür über ausreichende Ressourcen, solange die Fakultäten in ihren Strukturen des 20. Jahrhunderts aufgestellt bleiben, wie das gegenwärtig noch überwiegend der Fall ist. Wir haben versucht, ein Korpus zur Klassischen Moderne aufzubauen (https://kolimo.uni-goettingen.de/index.html). Es umfasst etwa 40.000 Texte von mehr als 2.000 Autoren der deutschen Literatur um 1900. Da eine Korpusliteraturwissenschaft (noch) nicht existiert, gibt es auch keine Infrastrukturen aus Datenbanken-Standards, Webspace, Urheberrecht und überhaupt einem Geist des Teilens von Ressourcen, um ein repräsentatives und balanciertes Korpus der Literatur um 1900 aufzubauen, also ein Korpus, das die verschiedenen Register, Domänen, Gattungen und Genre, Autorinnen und Autoren, Schulen und Richtungen historisch möglichst genau abbildet.[6] Vom Rechenzentrum bis zu den Metadatenstandards müssen jeweils erst einmal Lösungen formuliert, vorgeschlagen und erprobt werden, das aber in einem Fach, das selbst noch gar nicht weiß, ob Infrastrukturen für den Bau und Betrieb von Korpora überhaupt zu diesem Fach, hier der Neugermanistik, gehören. Texte von Kafka oder Thomas Mann sind zu einem erheblichen Teil für eine nicht nur buchzentrierte Forschung gar nicht zugänglich, jedenfalls nicht legal. Speicherformate und Metadaten sind nicht vereinheitlicht, wenn man etwa die Standards des *Deutschen Textarchivs* mit denen von *Gutenberg* und *TextGrid* vergleicht. Zitierfähige URLs für wissenschaftliches Bibliographieren müssen erst entwickelt und Datenzentren für geisteswissenschaftliche Daten erst etabliert werden. Das alles ist aufwändig und es genießt im Fach kein Ansehen.

Die Schwierigkeiten nehmen noch zu, wenn man das digital born-Material in den Blick nimmt. Denn nicht nur historisch in die Tiefe und kulturell in die Breite erlaubt die Digitalisierung die

[6] Herrmann, Berenike / Lauer, Gerhard: Korpusliteraturwissenschaft. In: Osnabrücker Beiträge zur Sprachtheorie 92 (2018), S. 127-156.

Transformation der geisteswissenschaftlichen Gegenstandsfelder. Erstmals können wir heute in Echtzeit verfolgen, wer, was, wo schreibt, teilt, kommentiert, und weiterschreibt. Remix ist das Stichwort, das die historisch einmaligen Möglichkeiten des Fortschreibens im digitalen Zeitalter umschreibt, das die Werkästhetik nur als einen Fall neben vielen anderen Verdichtungen und Auflösungen von Ausdrucksformen kennt. Nicht dass es E-Books und Games gibt, verändert das Gegenstandsfeld der Geisteswissenschaften so sehr, sondern die Möglichkeiten des Fort- und Umschreibens in fast alle Richtungen. Dafür verfügen die meisten geisteswissenschaftlichen Fächer noch nicht über ausreichende Infrastrukturen. Und die Ablehnung der digitalen Kulturwelt bzw. der Rückzug auf die Werkästhetik dominiert viele Fächer.

Ein Beispiel mag das Gesagte illustrieren: *Wattpad*. Hier schrieben bislang mehr als 40 Millionen Menschen in sehr vielen Sprachen dieser Welt Literatur, vor allem Romanzen und Fantasy. *Wattpad* wirbt damit, dass eine von drei jungen Frauen einen *Wattpad*-Account habe. Das dürfte eine Übertreibung sein, macht aber den Anspruch deutlich, dass es eine digitale Kultur gibt, von der kein Feuilleton Notiz nimmt und die doch für so viele Menschen ihre Kultur ist. Entscheidend ist hierbei, dass diese digitale Kultur beständig Daten erzeugt. Wir verfügen nicht nur über die Texte, sondern auch über die Daten der Nutzer. Wir wissen, welche Geschichten und Plots mehr Aufmerksamkeit auf sich ziehen als andere, was die nationalen und regionalen Präferenzen etwa von Hörbüchern sind, welche Altersgruppen welche Musik hören und welche nicht, wie sich Musik durch Streamingdienste verändert oder ob Frauen und Männer immer noch anderes lesen – fast. Denn diese Daten gehören zumeist den Unternehmen wie *Wattpad*. Die Extraktion von Daten über die ästhetischen Präferenzen der Gegenwart ist nur äußerst aufwändig zu erheben. Die Rechtslage ist, vorsichtig gesagt, unklar. Unklar damit, wie Daten geteilt werden und Ergebnisse überprüfbar gemacht werden können. Der Umfang der Daten passt nicht mehr in die etablierten Gefäße der philologischen Fächer. Von Tausenden, ja Millionen von Datenpunkten zu sprechen, ist keine Übertreibung, sondern das Problem. Für unsere Forschergruppe zum Lesen und Schreiben auf den sozialen Plattformen wie *Wattpad*, *Goodreads*, *Amazon Reviews*, müssen wir Infrastrukturen erst aufbauen, um zu erforschen, welche Literatur die heutige Gesellschaft schreibt, teilt und liest.

Anders gesagt, bedeuten die Veränderungen im Gegenstandsfeld der Geisteswissenschaften, dass sie nur sehr bedingt abgebildet werden können, einfach weil keine Infrastrukturen dafür vorhanden sind und der Aufbau solcher Infrastrukturen kein oder nur sehr wenig Anerkennung in den Fächern erbringt. Zitierfähige DOI sind ein wichtiger Schritt, um solche Forschung zu erlauben. Aber warum sich die Mühe machen, wenn es so viel einfacher ist, noch einen Walter Benjamin-Aufsatz zu schreiben. Anerkennungssysteme sind träge, so auch hier.

Zusammenfassend lässt sich zunächst nur der vorsichtige Schluss ziehen, dass die Transformation der geisteswissenschaftlichen Gegenstandsfelder in den nächsten Jahren nur sehr begrenzte Veränderungen der Infrastrukturen nach sich ziehen wird, eher in den kleinen Fächern, eher mit Blick auf die Erschließungsmöglichkeiten des kulturellen Erbes und eher in Fächern, die große Daten handhaben müssen wie die Linguistik oder Archäologie. Für sie macht die Digitalisierung des Gegenstandsfeldes einen Unterschied. Die anderen werden an der eher kanonischen Konstitution ihres Gegenstandes festhalten.

Für diejenigen, die genügend Neugierde und Forscherdrang spüren, ist eine Veränderung jedoch schon jetzt vielfach zu benennen. Sie wird zumeist und eher missverständlich unter Big

Data verhandelt. Genauer handelt es sich um eine Verschiebung weg vom Kanon, noch genauer um die Annäherung an die Grundgesamtheit, statistisch gesprochen. Während die Geisteswissenschaften – nicht alle, aber erhebliche Teile – stark am Kanon der wenigen, besonderen Werke ausgerichtet sind und von dort aus generalisieren, rückt vor allem mit dem digital born-Material nicht nur Kanon, nicht nur Stichproben, sondern die Grundgesamtheit dessen, was an Musik gehört oder an Literatur geschrieben wird, in Reichweite der Forschung. Damit ändert sich, was und wie generalisiert werden kann. Was der Roman des 19. Jahrhundert ist, das kann jetzt auch nahe an tatsächlich allen im 19. Jahrhundert gedruckten Romanen untersucht werden. Das ändert die Aussagenreichweite, die Exaktheit, Reliabilität, Replikation und ändert deren Validierung. Infrastrukturen müssen sich an diese veränderten Forschungsgegenstände anpassen, – das wäre zumindest die Forderung der Avantgarde der nicht etablierten Geisteswissenschaften.

3. Infrastrukturen und die Methoden der Geisteswissenschaften

Damit ist schon die zweite Veränderung, die der Methoden und der Methodik, implizit angesprochen. Der Kanon kann von Auge und Hand gelesen werden, andere Forschungsgegenstände des 21. Jahrhunderts brauchen Mal das Mikroskop oder mal das Teleskop, also ‚scalable reading'[7], Methoden, die den Abstand zum Text bzw. Objekt je nach Fragestellung wählen und dann entsprechende Instrumente nutzen. Solche Transformationen der Methodik lassen sich in einigen Fächern fraglos beobachten. Die Archäologie ist so selbstverständlich digitale Archäologie geworden, dass das Wort ‚digital' verschwunden ist, so wie sich das Fach Bioinformatik auflöst, weil es selbstverständlicher Teil der Biologie geworden ist. Computergestützte Explorationsverfahren wie LIDAR-Radartechnik, GIS-Verfahren für Ausgrabungen oder Wegeaufwandsberechnungen haben im Fach Archäologie so eminente Vorteile, weil sie eine Forschung ermöglichen, die ohne computergestützte Verfahren nicht möglich wäre, so dass diese digitalen Methoden zum Selbstverständnis eines Fachs dazu gehören. Ähnliches gilt für die Linguistik und Editorik. Auch hier lässt man den Zusatz ‚digital' weg, so selbstverständlich ist die Transformation der Methoden geworden. Dass die Linguistik mit *CLARIN* über eine eigene, disziplinäre Infrastruktur verfügt, ergibt sich aus der langen Tradition quantitativer und korpusbasierter Methoden, für die die Nutzung computergestützter Methoden nur die Fortsetzung ihrer bisherigen Forschungsmethodik mit anderen Mitteln ist. Dass das *Deutsche Archäologische Institut* über eigene digitale Infrastrukturen zur Handhabung seiner Daten aufgebaut hat, ist schiere Notwendigkeit für das Funktionieren dieser Forschung. Und Editionen etwa im Rahmen der *Union der deutschen Akademien* sind innerhalb einer Dekade ganz selbstverständlich digitale Editionen geworden.

Das gilt aber eben nur für einige Fächer, für andere nicht. In unserem Band *Die Rückkehr des Autors* von 1999 haben wir einen Beitrag des Pioniers der Autorenstylistik, John Burrows, abgedruckt. Unter allen Beiträgen blieb dieser Beitrag gänzlich unbeachtet, obwohl Burrows' Forschung gezeigt hat, dass Autorschaft nicht nur eine nachgetragene, soziale Konstruktion ist, sondern sich Autoren, Autorengruppen und Epochen anhand der Sprache unterscheiden lassen.[8] Während in den literaturwissenschaftlichen Fächern noch die Meinung dominiert,

[7] vgl. Mueller, Martin: https://scalablereading.northwestern.edu/.

[8] vgl. Burrows, John. F.: Computation into Criticism. A Study of Jane Austen's Novels and an Experiment in Method. Oxford: Clarendon Press 1987.

dass sich fiktionale und faktuale Texte nicht objektiv unterscheiden ließen, sondern auch dies nur eine soziale eingeübte gesellschaftliche Konvention sei, hat die computergestützte Forschung von Andrew Piper[9] beispielhaft gezeigt, dass es dazu mindestens gegenläufige Befunde gibt. Solche Befunde können Fächer wie die Anglistik problemlos ignorieren und tun das auch. Die Infrastrukturen, die hinter dieser Forschung stehen, – das sind vor allem die Entwicklung von Skriptsprachen Python und R und die Software-Plattform *Github* –, werden der Informatik zugeschlagen und spielen für die Fächer der philosophischen Fakultät keine größere Rolle. Die Entwicklungen zu einer Semantisierung der kulturellen Daten durch rdf und *IIIF* sind ohne Bedeutung für die Kunstgeschichte. Auch hier gilt, dass es erst dann einen Druck auf die Fächer gibt, sich digitalen Infrastrukturen als Aufgabe zu stellen, wenn die Ergebnisse nicht mehr ignoriert und die Ergebnisse nur auf den computergestützten Wegen zu erreichen sind.

Das ist gegenwärtig besonders deutlich im Feld des Machine Learning. Während das in den meisten Fächern vor allem kulturkritische Reflexe auslöst, hat Machine Learning einen rasant anwachsenden Einfluss auf ganz unterschiedliche Fächer. In Fraktur gedruckte Quellen sind auf einmal maschinenlesbar, damit erhebliche Teile der deutschsprachigen Kulturtraditionen. Muster in den Farben und Formen in großen Bildersammlungen lassen sich ermitteln und Zusammenhänge aufzeigen, die mit dem bloßen Auge nicht zu erkennen sind. Handschriften-Bestände wie etwa das globalhistorisch so spannende *Archiv der Basler Mission und Schweizerischen Handelsgesellschaft* können jetzt erschlossen werden und eröffnen Forschungsmöglichkeiten für globalhistorische Fragestellungen. Doch verlangt Machine Learning den Philologen und Kunsthistorikern eine steile Lernkurve ab, die Fachvertreter mit zunehmendem Etablierungsgrad nicht auf sich nehmen, sondern bestenfalls etwa an Digitalisierungszentren in Bibliotheken zu delegieren versuchen. Noch, denn es ist nicht auszuschließen, dass die computationellen Lernverfahren bald schon über einzelne Anwendungen hinaus etwa die Zusammenhänge der antiken Mittelmeerkultur oder die Wanderung buddhistischer Ideenwelt in einer Geschichte der *longue durée* aufzeigen können, die mit herkömmlichen Methoden nicht aufgezeigt werden können. Dann dürften auch die etablierten Fachvertreter ihr Methoden-Portfolio überdenken.

Die Verknüpfung von linguistischen Datenreihen mit DNA-Daten ist gegenwärtig ein Beispiel, wie Innovationen mit Datenbanken und computergestützten Modellierungen die Trennung in Natur- und Geisteswissenschaften aufheben und radikal andere Methoden in die Geisteswissenschaften einbringen. Infrastrukturen bis hin zur Gründung eines eigenen *Max- Planck-Instituts für Menschheitsgeschichte* ermöglichen eine Forschung, die bis vor kurzem für unmöglich gehalten wurde, eben Menschheitsgeschichte zu betreiben, also die Zusammenhänge von paläoanthropologischen DNA-Daten mit der Verbreitung von Sprachen und deren Strukturen zu verstehen. Für Fächer wie die Geschichte oder die Sprachwissenschaft sind die Ergebnisse von Svante Pääbo, Simon Greenhill oder Patrick Geary mindestens eine Irritation und können nur soweit als Randphänomen außen vorgehalten werden, wie sie eher die Ur- und Frühgeschichte betreffen und nicht die Frage klären können, wie es zum Aufstieg des Nationalsozialismus gekommen ist. Ansonsten zeigen sie, dass ganz andere, hier naturwissenschaftliche Ansätze zum Methodenset der Geisteswissenschaften zählen können und eine solche Forschung völlig neue Infrastrukturen benötigt, eben auch DNA-Labore.

[9] vgl. Piper, Andrew: Fictionality. In: Journal of Cultural Analytics (2016).

Anders gewendet gilt auch im Feld der Methodenentwicklung, dass aus neuen methodischen Möglichkeiten nicht zwingend Veränderungen der Fächer und ihrer Infrastrukturen folgen. Während die Linguistik über aufwändige Datenbanken zur Kartographierung der Sprachen der Welt verfügt, kennen so etwas die literaturwissenschaftlichen Fächer nicht, trotz deren emphatischer Rede von der Weltliteratur. Die Rekonstruktion der Namen der 6 Millionen ermordeten Juden braucht eine aufwändige Infrastruktur, *EHRI- European Holocaust Research Infrastructure*, um Name Entities aus Deportationslisten und Steuerakten zusammenzuführen, andere Teile der Geschichtswissenschaft aber nicht. Die Wanderungen Ideen des Aristotelismus durch die verschiedenen Orden, Klöster und entstehenden Universitäten des Mittelalters nachzuverfolgen, ist jetzt möglich, auch die Korrelationen zu entdecken, die vielleicht mit der Geographie, Siedlungsdichte oder dem römischen Straßennetz zusammenhängen könnten. Für andere Bereiche der Philosophie- und Ideengeschichte werden solche Fragen nicht gestellt. Kurz: Neue Formen der Historiographie sind möglich, aber das sind bewusste Entscheidungen, keine Notwendigkeiten, die sich aus der Digitalisierung ergeben würden.

Ob man sich für solche, infrastrukturell aufwändigen Ansätze entscheidet, hängt an zwei Abwägungen. Das eine ist die Bereitschaft, neben den historischen-hermeneutischen Ansätzen auch solche der Formalisierung von Problemen zuzulassen. Praktisch alle computergestützten Ansätze verlangen, dass Probleme in möglichst exakte Kategorien modelliert werden müssen, um von Maschinen verarbeitet werden zu können. Die Vektorisierung und Faktorisierung, die Bestimmung von Variablen und der Ausschluss von konfundierenden Variablen, die Verlässlichkeit von Metadaten und deren Standardisierung sagen immer eines: dass die vertrauten historisch-hermeneutischen Methoden um solche der exakten Wissenschaften zu erweitern oder gar in diese exakten Begriffe und Konzepte zu überführen sind. Computer verstehen nichts und können auch nicht lesen oder beobachten. Ihre Leistungen hängen an der Genauigkeit und methodische Strenge, die Menschen durch das experimentelle Design in die Untersuchung einbringen. Computer sind nur so exakt wie die Wissenschaftler es sind, die die Forschung betreiben. Und dafür sind nun eher formale und eher quantitative Denkweisen notwendig, die in den Geisteswissenschaften nicht oder nur etwa in Fächern wie etwa der Linguistik eingeübt werden. Die Lernkurve ist entsprechend steil, wenn jemand Texte mit der Statistiksoftware R untersuchen möchte. Ohne Modellierung von Fragestellungen ist hier nichts zu erreichen und ich übertreibe nicht, wenn ich die Modellierung von Fragestellungen das Schibboleth der Geisteswissenschaft des 21. Jahrhunderts nenne. An ihm hängt wesentlich, ob die digitale Transformation der Geisteswissenschaften gelingt.

Eng damit zusammen hängt dann der für die Infrastruktur so wesentliche Punkt der Stellen. Wollen die Geisteswissenschaften mehr sein als nur ausstehende Beobachter, die warnend den Finger erheben angesichts der fortschreitenden Digitalisierung, wird es darauf ankommen, ob die Fächer Stellen zur computergestützten Modellierung ihrer Arbeit einrichten oder nicht. Denn nicht Datenbanktechniken oder Datenstandards, Rechnerkapazitäten oder Statistik sind so schwierig in den Fächern zu verankern, sondern die Umschichtung von Stellen und die Gewöhnung daran, dass nicht der Einzelwissenschaftler, sondern Teams die Geisteswissenschaften des 21. Jahrhunderts ausmachen. Man braucht eine Machine Learnerin und vielleicht noch einen Datenbank-Spezialisten, schreibt ein Paper zu Robert Musil mit vier anderen Kollegen, die nicht alle Germanistik als Fachhintergrund haben, sondern etwa auch Informatik, man braucht eine andere Infrastruktur der Stellen ähnlich der in den Naturwissenschaften. Damit werden auch Geisteswissenschaften teurer. Politikwissenschaften suchen *Twitter*-Ana-

lysten und die Archäologen brauchen für ihre Forschung Radarspezialisten. Der Erfolg so unterschiedlicher Projekte wie die Forschung zur sefardischen Polemikliteratur in der Frühen Neuzeit oder die Forschung zur Verbreitung der St. Gallener Hymnik im Europa des Mittelalters hängt daran, ob man genügend Fachleute für solche Vorhaben findet, die eben nicht mehr nur aus dem jeweils eigenen Fach wie Iberoromanistik oder Musikwissenschaft kommen. Dafür braucht es Stellen, genauer eine andere Stellenstruktur, die Technikerstellen nicht nur als Hilfswissenschaftler behandelt. Sie sind der Schlüssel für die Geisteswissenschaften des 21. Jahrhunderts. Die Modellierung von Forschungsfragen braucht andere Forschungsteams als es noch etwa Grimms Wörterbuch gebraucht hat. Die Geisteswissenschaften des 21. Jahrhunderts werden vielfach nicht mehr nur Geisteswissenschaften bleiben können, wenn sie den Anforderungen computergestützter Modellierung entsprechen wollen und das hängt letztlich an den Stellenprofilen. Entweder setzen die philosophischen Fakultäten auf die Einzelwissenschaftler oder auf Teams. Und das bedeutet, sie müssen die Fächer anders als bisher vernetzen, um arbeitsfähige Gruppen aufzubauen.

4. Infrastrukturen und Umwelten der Geisteswissenschaften

Veränderungen der Fächer durch die Digitalisierung werden nicht durch die Digitalisierung der Gegenstandsfelder noch durch die computationelle Modellierung und Formalisierung von Methoden jeweils alleine nahegelegt. Das gilt, drittens, auch für die Veränderungen der Akteure und Umwelten der Geisteswissenschaften. Es ist schon deutlich geworden, dass eine, wenn nicht die wesentliche Veränderung der digitalen Geisteswissenschaften darin liegt, dass sie eher von Forschergruppen umgesetzt werden. Das bedeutet nicht, dass es nicht auf Einzelarbeit ankäme und das einsame Nachdenken bedeutungslose geworden wäre. Auch im *CERN* kommt es darauf immer noch an, so auch in der Teamarbeit in den Geisteswissenschaften des 21. Jahrhunderts. Aber Korpusaufbau und Datenbanktechnik, Textmining und Archivkenntnisse, Machine Learning und Edition meinen verschiedene Kompetenzen und sind selten in einer Person vereint. Daher ist die Arbeit in Gruppen eine der potentiellen Veränderungen, die nicht wenig Irritation in den gerne am Genie-Paradigma angelehnten Geisteswissenschaften auslöst.

Mit der Veränderung der Akteure meine ich aber nicht nur die Betonung von Forschergruppen, sondern auch die andere Vernetzung, die erst neue Forschung ermöglicht. Für die Geisteswissenschaften sind besonders Bibliotheken nicht nur ein Ort, wo man Bücher findet und sie in Ruhe liest, sie sind ein Forschungspartner und das meint mehr als nur die Infrastruktur von Digitalisierungszentren, die zumeist an Bibliotheken angesiedelt sind. Es meint auch die gemeinsame Entwicklung von Datenstandards und Austauschformaten bzw. -technologien wie *IIIF*, es meint das gemeinsame Schreiben von Forschungsanträgen, weil nur zusammen mit Bibliotheken ein Projekt vom Papyrus in einer Sammlung wie in der UB Basel bis zur Machine Learning-gestützten Untersuchung der Zusammenhänge dieser Dokumente gelingen kann.

Eine weitere Veränderung gibt es auch auf der Seite der Publikation der Ergebnisse. Ergebnisse sind nicht mehr nur der Aufsatz oder das Buch, sondern auch der Code oder die Datenbank. Publikationen auf *Github*, Preprints auf *ArXiv.org*, Datenbanken im Netz, sind Teil der Publikation, wenn Replikation und Reliabilität ernstgenommen werden sollen. Open Science ist daher die Kurzformel für einen Umbau, der vor allem in den etablierten Geisteswissenschaften auf

Ablehnung stößt. Für die nicht-etablierten Geisteswissenschaften sind hingegen solche Publikationsformen essentiell, weil die Leistungen auch in zitierfähigen Editionen, Datenbanken, rdf-Ontologien, Trainingsets oder Scripts bestehen können. Mehr noch entsteht ein Mehrwert daraus, wenn Daten nachgenutzt werden können, rekombiniert und mit anderen Befunden verglichen werden können. Für unser Projekt zur Geschichte des Tagebuchsschreibens in Zeiten des Kriegs bringt ein einzelnes Tagebuch wenig; viel dagegen, wenn etwa Anne Franks Tagebuch mit anderen Tagebüchern verglichen werden kann, und damit erst Eigentümlichkeit des Schreibens Jugendlicher unter solchen Belastungen wie im Versteck in der Prinsengracht gesehen und Ähnlichkeiten bestimmt werden können. Die korpusbasierte Arbeit entsteht eben erst, wenn wir bereit sind die Daten zu teilen, ob das Editionen oder einfach nur Textfiles sind, und dann über Portale mit der Community nicht nur für uns behalten. In Fächern wie der Archäologie ist der Mehrwert längst evident, wenn Sammlungen wie die des *British Museum* ihre Bestände über das Netz teilen. Eine solche Sammlung wie die des *British Museums* über das Internet zu teilen ist Infrastruktur in dem elementaren Sinn, um den es mir hier geht.

Schließlich ändern sich auch Umwelten. Es ist keine Seltenheit mehr, dass Mathematiker sich mit geisteswissenschaftlichen Fragestellungen befassen, das haben schon Thomas de Morgan oder Corwin Mendenhall getan, als sie sich mit der Frage der Urheberschaft der Werke Shakespeares oder der Autorschaft der Briefe des Paulus befassten.[10] Das war im 19. Jahrhundert und vielleicht vielfach nur Liebhaberei. Heute nutzen Mathematiker statistische Modelle zur Überprüfung der Pollyanna-Hypothese oder zur Untersuchung der Plotstruktur von Romanen. Ganz andere Ansätze wie die DNA-Analyse greifen in die Bereiche der Geschichtswissenschaft oder Linguistik über. Die Fakultätsgrenzen des 20. Jahrhunderts lösen sich gleich an mehreren Stellen auf. Taxonomische Modelle wie Splitstrees werden für die Analyse von Märchenmotiven genutzt, eine Fragestellung in den Geisteswissenschaften, die schon mehr als hundert Jahre alt ist und von der Finnischen Märchenschule im *Aarne- Thompson-Uther-Index* installiert wurde.[11] Jetzt sind solche taxonomischen Modelle der Biologie auf einmal mitten in den Geisteswissenschaften zu finden. Noch findet eine solche Forschung am Rande der Mehrheits-Geisteswissenschaften statt, etwa in der *Max-Planck-Gesellschaft*.

Die Liste der Beispiele lässt sich leicht verlängern. Deutlich wird dabei, dass die Geisteswissenschaften des 21. Jahrhundert mindestens nicht notwendig die des 20. Jahrhunderts sein müssen. Im Gegenteil sind andere Arbeitsformen längst erprobt, andere Formen der Zusammenarbeit vielfach umgesetzt, andere Publikationswege mehr als nur ein Versprechen und andere Grenzziehungen im Streit der Fakultäten möglich.

5. Geisteswissenschaften für das 21. Jahrhundert

Keine der Veränderungen alleine drängt eine Transformation der Geisteswissenschaften auf und legt eine Neujustierung der bisherigen Infrastrukturen dafür nahe. Veränderungen in den Wissenschaften sind eben von vielen Faktoren abhängig, vor allem in welchem Verhältnis Aufwand und Ertrag für einzelne Wissenschaften zueinander stehen. Zeit in das Erlernen statisti-

[10] vgl. Gryzbek, Peter (Hg.): Contributions to the Science of Text and Language. Word Length Studies and Related Issues. Dordrecht: Springer 2006.
[11] vgl. Tehrani, Jamshid: The Phylogeny of Little Red Riding Hood. In: PLoSONE 8(11) (2013), e78871.

scher Methoden mit R zu investieren, wird nur der aufbringen, der sich einen im Fach angesehenen Gewinn davon versprechen kann. Auf *ArXiv.org* wird nur die publizieren, die weiß, dass ihre Arbeit dort auch von Fachkollegen gelesen wird. Warum die Zeit mit Studenten und Kollegen verbringen, wenn ich am besten in Ruhe mein Buch schreiben kann, gerade wenn der Horizont der Fachwahrnehmung eher im Feuilleton denn im Fach zu finden ist. Den *Leibniz-Preis* vergibt die *DFG* in den Geisteswissenschaften selten an risikobehaftete Forschung. Es hat schlicht viele Vorteile, der etablierten Wissenschaftskonvention zu folgen. Gegebenenfalls sind die Sanktionen scharf, die Fächer gegenüber Minderheiten erlassen können. Es gibt also nur wenige Gründe, dass sich etwas ändern wird.

Aber vielleicht sind unter den Geisteswissenschaftlern und -wissenschaftlerinnen auch genügend, die allen Widerständen zum Trotz gerade deshalb Wissenschaften als Beruf gewählt haben, weil sie auch die Umwege oder wenig begangenen Wege einschlagen wollen und dabei das Risiko eingehen, dass dies auch Holzwege sein können. Wer die breiten Straßen verlässt, macht Entdeckungen und die brauchen dann andere Infrastrukturen. Wenn also Gegenstand des Fachs nicht nur der Kanon ist, sondern eher die Population, wenn die Methoden auch quantitative und experimentelle Ansätze umfassen, wenn andere Formen des Arbeitens, vor allem solche des Zusammenarbeitens und das auch über Fakultätsgrenzen hinweg und mit anderen Akteuren wie den Bibliotheken die eigene Neugierde anleiten, weil man voneinander lernen kann für die jeweils eigene Forschung, dann ändern sich auch Infrastrukturen.

Diese Infrastrukturen haben die folgenden Merkmale:

- Sie ermöglichen den flexiblen, je an die Fragestellung angepassten Aufbau von Korpora.
- Sie bieten Schnittstellen, zur Weiternutzung der Korpora.
- Sie erlauben den Export für verschiedene Miningverfahren, dem Rechnen auf virtuellen Maschinen etwa.
- Sie befördern die Kollaboration etwa durch Schnittstellen für Zusammenarbeit mit Bibliotheken, über *Gitlab* oder Veröffentlichungen über *Github*.
- Sie erlauben persistente Verlinkungen und stabile Versionierungen von Daten.
- Sie sichern sensible Daten.
- Und sie ermöglichen es, dass noch wenig erprobte Ansätze Pilotstudien fahren können.

Aber entscheidend ist das alles nicht. Entscheidend ist erstens, ob die philosophischen Fakultäten andere, eher an Forschergruppen orientierte Stellenstrukturen etablieren oder nicht. Zweitens ist wesentlich, ob die eher formalen Modellierungen von geisteswissenschaftlichen Fragestellungen nicht zuletzt auch in der Lehre der Fächer verankert werden.

Ich weiß nicht, ob in etwa fünf Jahren Infrastrukturen für die vermutlich auch dann noch kleine Gruppe von experimentierfreudigen Geisteswissenschaftlern und -wissenschaftlerinnen hinreichend etabliert sein werden, damit nicht jeder Schritt einer Grundsatzentscheidung gleich kommt. In einzelnen Fächern ist das ja schon gelungen, in meinem Fach der Literaturwissenschaft ist das noch die Zukunft. Und die hat bekanntlich den Vor- oder Nachteil unbekannt zu sein.

Francis Bacon hatte vor mehr als 400 Jahren verschiedene Idola unterschieden, die uns Menschen daran hindern, Erkenntnisse zu gewinnen, solche der eher unbewussten Täuschungen, der Dogmen, der Gewohnheit, aber auch des Verstandes. Es sind wohl letztere, die uns davon

abhalten, in unseren Philosophischen Fakultäten überhaupt nur die Diskussion zu beginnen, ob andere Infrastrukturen für die Geisteswissenschaften des 21. Jahrhunderts möglich sind. Mein Beitrag ist ein schwaches Plädoyer dafür diese Diskussion um die Geisteswissenschaften des 21. Jahrhunderts zu beginnen, bevor das stille Sterben der Geisteswissenschaft diese Fächer aufgelöst haben wird.

Literaturverzeichnis

Bacon, Francis: The Advancement of Learning. In: Francis Bacon. The Major Works. Edited with an Introduction and Notes by Brian Vickers. Oxford: Oxford University Press 1996 [1605], S. 120-299.

Burrows, John. F.: Computation into Criticism. A Study of Jane Austen's Novels and an Experiment in Method. Oxford: Clarendon Press 1987.

Diemer, Alwin: Art. Geisteswissenschaften. In: Historisches Wörterbuch der Philosophie 3 (1974), S. 211-215.

Gaukroger, Stephen: Francis Bacon and the Transformation of Early-Modern Philosophy. Cambridge: Cambridge University Press 2010.

Gryzbek, Peter (Hg.): Contributions to the Science of Text and Language. Word Length Studies and Related Issues. Dordrecht: Springer 2006.

Herrmann, Berenike / Lauer, Gerhard: Korpusliteraturwissenschaft. In: Osnabrücker Beiträge zur Sprachtheorie 92 (2018), S. 127-156.

Holliger, David. A.: The Humanities and the Dynamics of Inclusion since World War II. Baltimore: Johns Hopkins University Press 2006.

Lauer, Gerhard (Hg.): Wilhelm von Humboldt. Schriften zur Bildung. Stuttgart: Reclam 2017.

Mueller, Martin: https://scalablereading.northwestern.edu/ (Zugriffsdatum: 05.12.2018).

Piper, Andrew: Fictionality. In: Journal of Cultural Analytics (2016), DOI: 10.31235/osf.io/93mdj bzw. http://culturalanalytics.org/2016/12/fictionality/ (Zugriffsdatum: 05.12.2018).

Rebenich, Stefan: Theodor Mommsen und Adolf Harnack. Wissenschaft und Politik des ausgehenden 19. Jahrhunderts. Berlin: de Gruyter 2012.

Tehrani, Jamshid: The Phylogeny of Little Red Riding Hood. In: PLoSONE 8(11) (2013), e78871. https://doi.org/10.1371/journal.pone.0078871 (Zugriffsdatum: 05.12.2018).

II. Beispiele aus den Fächern

Datenerhebung als epistemologisches Labor – Überlegungen am Beispiel der virtuellen Forschungsumgebung *Revistas culturales 2.0*

Hanno Ehrlicher, Jörg Lehmann (Tübingen)

1. Die Forschungsumgebung *Revistas culturales 2.0* und ihr Potential als epistemologisches Labor für die Kulturzeitschriftenforschung

Seit 2014 existiert die virtuelle Forschungsumgebung *Revistas culturales 2.0*. Hier erkundet ein internationales Forschungsnetzwerk digitalisierte spanischsprachige Kulturzeitschriften. Die Digitalisate werden im Rahmen einer Kooperationsvereinbarung vom Ibero-Amerikanischen Institut Preußischer Kulturbesitz (IAI) bereitgestellt und über eine Schnittstelle in die Forschungsumgebung importiert. Derzeit können rund 112 Zeitschriftentitel bearbeitet werden, das Korpus an Digitalisaten wächst aber parallel zur Digitalisierung des Kulturzeitschriftenbestands des IAIs kontinuierlich an und ließe sich perspektivisch durch weitere Kooperationen mit den relevanten Bibliotheken aus dem spanischsprachigen Raum auch noch systematisch erweitern. Ein Beispiel hierfür wäre die Madrider *Biblioteca Nacional de España*, die in den letzten Jahrzehnten besonders aktiv bei der Digitalisierung ihrer hemerographischen Bestände war.[1] Ziel der *crowdsourcing*-Herangehensweise ist die Erkundung der historischen Vernetzung der spanischsprachigen Kulturzeitschriften, die als Medium einer Modernisierungsdynamik gesehen werden. Das Portal leistet damit zweierlei: Erstens die Überwindung der nationalen Logik der Bibliotheken und Archive und die damit einhergehende Zersplitterung der Forschung, die für den spanischsprachigen Kulturraum besonders kennzeichnend ist, nicht nur was den transatlantischen Austausch zwischen Spanien und den lateinamerikanischen Staaten betrifft, sondern insbesondere auch bezüglich des länderübergreifenden Austausch innerhalb des Kulturraums, der einem bereits wenige Jahrzehnte nach den staatlichen Unabhängigkeitsprozessen des 19. Jahrhunderts verbreiteten diagnostischen Bonmot zufolge auch gerne als „Estados *des*unidos de la América del Sur“ bezeichnet wird.[2] Die Forschungsumgebung bietet den Forscher*innen weltweit jedoch nicht nur den Zugang zu Digitalisaten und ihren Metadaten, sondern ermöglicht ihnen zugleich auch, und das ist der entscheidende zweite Mehrwert der Forschungsumgebung , ihren individuellen Fragestellungen nachzugehen und die jeweils für ihr eigenes Forschungsprojekt relevanten Daten zu erheben. Neben

[1] Die „Hemeroteca Digital“ der BNE bietet derzeit über 2.000 Zeitschriften- oder Zeitungstitel an, die weit überwiegende Mehrheit davon im Open Access. Vgl. http://www.bne.es/es/Catalogos/HemerotecaDigital/docs/tabla_listadocompleto.pdf.

[2] Der Ausdruck „estados des-unidos“ wurde beispielsweise schon vom chilenischen Intellektuellen Francisco Bilbao 1856 verwendet, im nachträglichen Kommentar zu einem in Paris gehaltenen Vortrag „Iniciativa de América. Idea de un Congreso Federal des las Repúblicas“, wo der Autor die alte Bolivarsche Idee einer Föderation der Lateinamerikanischen Staaten aufgreift. Im „Post-Dictum“erläutert er seine Zielsetzung dabei wie folgt: „La idea de la Confederación de la América del Sur, propuesta un día por Bolívar, intentada después por un Congreso de plenipotenciarios de algunas de las Repúblicas, y reunido en Lima, no ha producido los resultados que debían esperarse. Los Estados han permanecido *Des-Unidos.“* Bilbao, Francisco: Post-Dictum. In: Manuel Bilbao (Hg.): Obras completas de Francisco Bilbao, Vol. 1, Buenos Aires: Imprenta de Buenos Aires 1866, S. 285–304, hier S. 285.

den Image-Dateien der digitalisierten Kulturzeitschriften werden nämlich auch die vom IAI bereitgestellten Metadaten im XML-Standard METS/MODS importiert (Metadata Encoding & Transmission Standard bzw. Metadata Object Description Schema). Das Portal ermöglicht dem Forscherkollektiv durch vorstrukturierte Annotationsformulare die systematische Anreichung dieser Metadaten (Abb. 1 und 2).

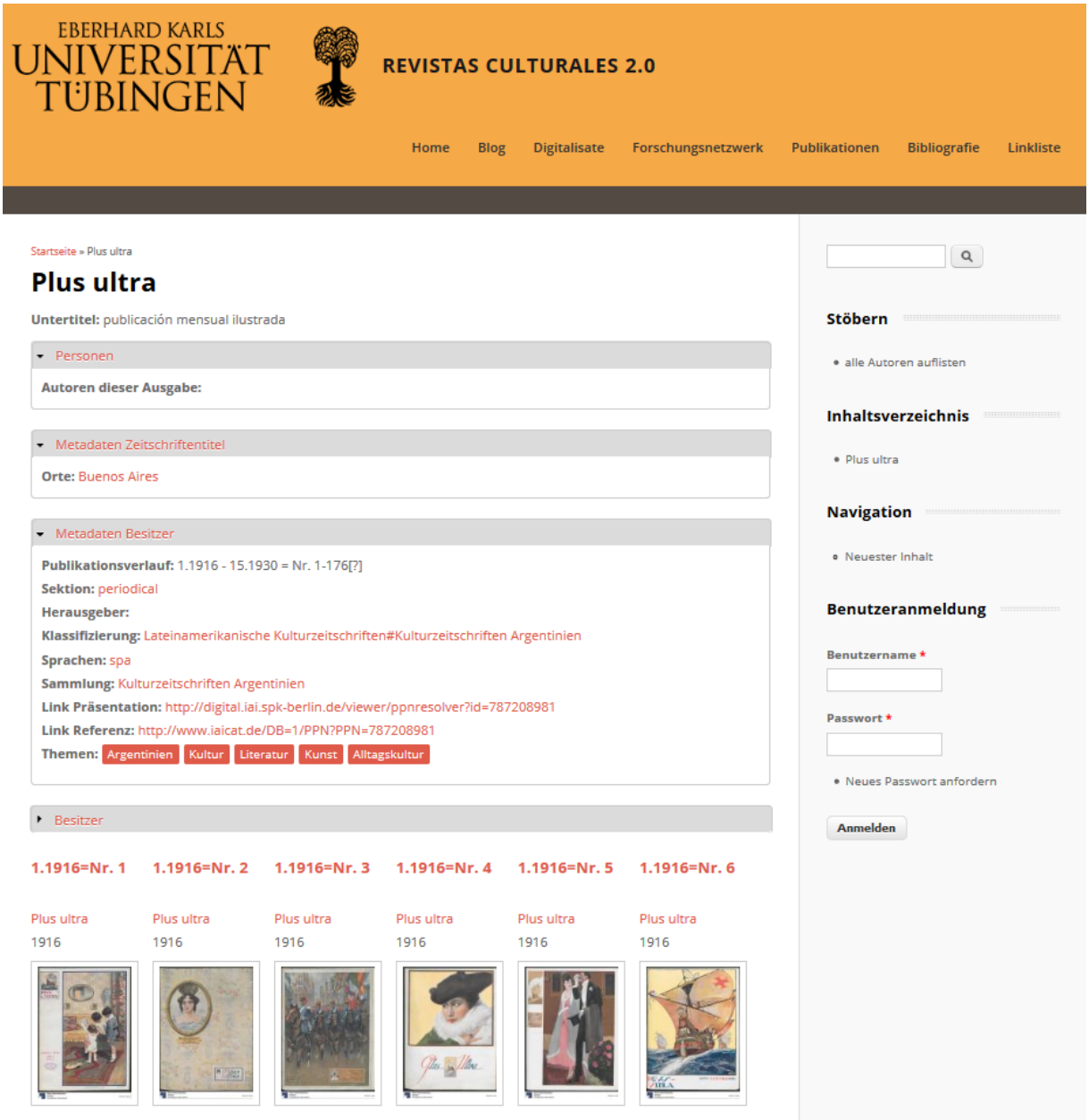

Abb. 1: Screenshot der virtuellen Forschungsumgebung Revistas Culturales 2.0 mitsamt den bibliothekarischen Metadaten; gewähltes Beispiel: Die Kulturzeitschrift „Plus Ultra“

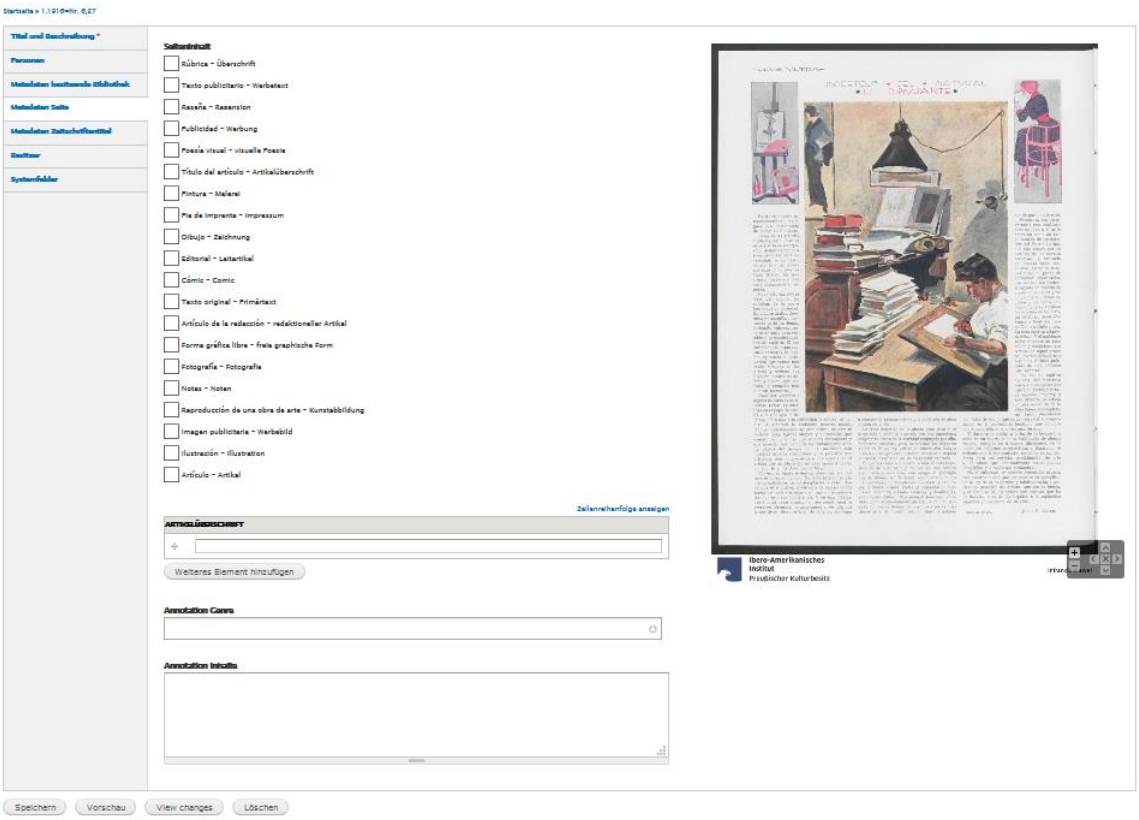

Abb. 2: Screenshot der Arbeitsumgebung in der virtuellen Forschungsumgebung Revistas Culturales 2.0; eine einzelne Seite der Kulturzeitschrift „Plus Ultra“

Wie die beiden Screenshots zeigen, werden den Forscher*innen zunächst die vom IAI erhobenen Metadaten präsentiert. In den Feldern der Forschungsumgebung können die Forscher*innen dann weitere Daten selbst erheben, indem sie vorgegebene Felder anklicken oder ausfüllen (z.B. Artikelüberschriften) oder thesaurierend etwa neue Genres hinzufügen, oder sie können entsprechend ihrer individuellen Fragestellungen Annotationen erstellen. Nachdem eine Zeitschrift durchgearbeitet wurde, können sämtliche erhobenen Daten im csv-Format exportiert werden.

Bereits hier zeigt sich ein grundsätzlicher Unterschied zur Vorgehensweise der klassischen Literaturwissenschaft. Diese nimmt einen einzelnen oder eine Gruppe von Texten gerne als Grundlage der Erforschung und betrachtet ihre ‚Daten‘ als dem epistemischen Prozess vorgängig. Die Texte selbst sowie die Angaben zu Autor, Verlag, Erscheinungsjahr usf. bilden die Daten der Literaturwissenschaft, sie werden als dem Erkenntnisprozess vorausgehend und ‚gegeben‘, im Wortsinne also „datum“ angesehen. Auf diese Art und Weise verfahren im Grunde heute viele Big-Data-Analysen[3] und tendenziell auch Franco Morettis Ansätze zum „Distant Reading“, sofern sie zwar von einem quantitativ erweiterten Archiv von Literatur im Weltmaßstab ausgehen und das ‚Lesen‘ dabei zunehmend den Maschinen überlassen, die von

[3] Dass „Big Data“ dem epistemischen Prozess meist vorgängig sind, zeigte sich im K-PLEX-Project. Vgl. dazu Lehmann, Jörg / Huber, Elisabeth / Stodulka, Thomas: H2020 Project K-PLEX: WP4 Report on Data, Knowledge Organisation and Epistemics, Research Report, Berlin: Freie Universtät Berlin 2018, S. 75. HAL Id: <hal01761214>. https://hal.archives-ouvertes.fr/hal-01761214 (Zugriffsdatum: 20.11.2018).

den Nationalphilologien ererbten Kategorienbildungen dabei jedoch als quasi-empirische Befunde ohne weitere Problematisierung übernehmen.[4] Im Gegensatz dazu müssen in der virtuellen Forschungsumgebung *Revistas Culturales 2.0* jene Daten, die die Grundlage der Auswertung bilden, überhaupt erst erhoben werden. Hier zeigt sich, dass Daten sowohl als Input für einen Forschungsprozess als auch als Output dieses Prozesses angesehen werden können. Letzteres ist dann der Fall, wenn sie Ergebnis einer Erhebung im Rahmen eines epistemologischen Prozesses sind. Diese Beobachtung ist in den Sozialwissenschaften selbstverständlich. Hier werden nicht nur ökonomische und sozialstrukturelle Daten verarbeitet, wie sie in amtlichen Statistiken bereitgestellt werden (DeStatis, Eurostat, UNO), sondern ein bedeutender Teil der Daten muss überhaupt erst selbst erhoben werden, vor allem durch Fragebögen. Eine solche Herangehensweise – und die damit verbundene ausgefeilte Methodologie – ist in den Geisteswissenschaften jedoch alles andere als selbstverständlich. Dieser fehlenden Verankerung in den geisteswissenschaftlichen Fächern ist es auch geschuldet, dass Datenerhebungsangebote wie das auf unserer Forschungsumgebung offerierte bisher eher zögerliche Annahme in der internationalen Forschercommunity gefunden haben. Literaturwissenschaftler etwa müssen sich Methoden der Klassifikation und Evaluation erst aneignen. Epistemologisch neu für die Geisteswissenschaften ist die mit der formalisierten Klassifikationsleistung einhergehende De-Flexibilisierung, denn Klassifikationen, die formal stabil sein müssen, legen im Gegensatz zu Interpretationen fest und müssen bei einem Wechsel der interpretatorischen Hypothesen gegebenenfalls nachträglich entsprechend zeitaufwändig angepasst werden. Die epistemische Herausforderung geht zudem mit dem sozialpragmatischen Hindernis fehlender Gratifikation des betriebenen Aufwands einher, denn die für die Datenerhebung zentralen Klassifikationsleistungen sind äußerst zeit- und damit ressourcenintensiv. Sie werden im Prozess der kognitiven Wertschöpfung in den klassischen Interpretationsverfahren der Hermeneutik, aber auch durch deren post-hermeneutischen methodischen Widersacher als solche aber praktisch überhaupt nicht valorisiert, sondern verschwinden ganz hinter den für die ‚Ewigkeit' der Nachwelt formulierten, argumentativ als Thesen begründeten Erkenntnissen. Man muss sich nur die Wochen und Monate vor Augen führen, die für die Datensammlung in unserem Projekt aufgewendet werden, und die sich konkret in den verbrauchten Personenmonaten ausdrücken lassen, um das eklatante Missverhältnis zwischen praktischer Bedeutung der Datenerhebung und der Gefahr fehlender Valorisierung einzuschätzen, eine Gefahr, der sich nicht ins Projekt eingebundene Forscher*innen aus dem Umfeld der traditionellen Philologien oder Kulturwissenschaft nicht ohne Weiteres aussetzen werden. Ein umfassendes Verständnis für die Datenerhebung als einem substantiellen und kostspieligen Teil des epistemologischen Prozesses scheint in den Geisteswissenschaften noch nicht angekommen zu sein und kann auch erst dann zum Bestand der habitualisierten Methoden werden, wenn Datenerhebung auch in den Evaluations- und Gratifikationssystemen der beteiligten Fächer entsprechende Berücksichtigung finden wird. Ein erster Schritt dazu soll die hier unternommene und

[4] Da Franco Moretti ein methodisch äußerst flexibler Wissenschaftler ist, der zu einer ständigen Falsifikation der eigenen Vorannahmen bereit ist, gilt es hier präzise zu sein. Zu Beginn der Rede vom *Distant Reading* war Moretti zunächst noch von einer „sort of cosmic and inevitable divison of labour" ausgegangen, bei welcher der aus der Distanz operierende Literaturwissenschaftler in einem „second hand criticism" Ergebnisse der nationalphilologischen Forschung und damit auch deren Kategorisierungen ungeprüft als Ausgangsfakten einsetzen und bis auf Weiteres blind verifizieren muss (Moretti, Franco: Conjectures on World Literature. In: New Left Review 1 (2000), S. 54–68, hier bes. S.66). Hingegen betont Moretti in jüngster Zeit umgekehrt das epistemische Verfremdungspotential der Methode, das auch und gerade zu einer Infragestellung der zunächst vorausgesetzten Kategorien durch die Prozesse der Operationalisierung führe: Moretti, Franco: Operationalizing. Or, the Function of Measurement in Literary Theory. In: New Left Review 84 (2013), S. 103–119, sowie online im Rahmen der Pamphlets des Stanforder Literary Labs: https://litlab.stanford.edu/LiteraryLabPamphlet6.pdf (Zugriffsdatum: 20.11.2018).

an konkreten Beispielen durchgeführte methologische Reflexion darstellen, die eben aufgrund des eminent praxisbezogenen Charakters der Datenerhebung naturgemäß praxeologischen Zuschnitt haben muss. Im Grunde greifen wir mit diesen Überlegungen dabei die epistemische Zwischenebene auf, die z.B. in den von Peer Trilcke und Frank Fischer jüngst vorgenommenen interessanten Reflexionen zur „Praxeologie der Digital Literary Studies und ihren epistemischen Dingen“[5] keine eigene Berücksichtigung fand. Diese epistemische Zwischenebene ist im Horizont unserer Forschungspraxis aber gerade entscheidend, und sie dürfte auch generell von großer Bedeutung sein für eine erfolgreiche oder doch zumindest verbesserte Vermittlung zwischen den ans Textmedium gebundenen „epistemischen Dingen“, die die klassische Literaturwissenschaft behandelt, und den codebasierten, die von den neuen digitalen Literaturwissenschaften modelliert und analysiert werden. Diese allgemeine Problemstellung soll nun im Folgenden anhand von konkreten Beispielen genauer erläutert werden. Sie wurden aus dem eng mit der virtuellen Forschungsumgebung *Revistas Culturales 2.0* verbundenen DFG-Projekt „Literarische Modernisierungsprozesse und transnationale Netzwerkbildung im Medium der Kulturzeitschrift“ gewonnen.

2. Metadatenerhebung als epistemische Herausforderung

In diesem nunmehr ca. in der Hälfte ihres Förderzeitraums befindlichen Forschungsprojekt wird ein Korpus von fast 60 digitalisierten Zeitschriften aus dem spanischsprachigen Kulturraum untersucht, wobei transnationale Akteursnetzwerke ebenso im Zentrum stehen wie das Netzwerk aus Textformen (Gattungen und Texttypen), das eine transnationale Übersetzung und einen Kulturtransfer literarischer Formen leistet. Die kulturübergreifende Anlage dieses Zeitschriftenkorpus lässt sich auf einer kartographischen Visualisierung leicht veranschaulichen.

[5] Trilcke, Peer / Fischer, Frank: Literaturwissenschaft als Hackathon. Zur Praxeologie der Digital Literary Studies und ihren epistemischen Dingen. In: Martin Huber / Sybille Krämer (Hg.): Wie Digitalität die Geisteswissenschaften verändert: Neue Forschungsgegenstände und Methoden, 2018 (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 3). text/html Format. DOI: 10.17175/sb003_003 (Zugriffsdatum: 20.11.2018).

Abb. 3: Visualisierung der räumlichen Verteilung der fast 60 im DFG-Projekt „Literarische Modernisierungsprozesse und transnationale Netzwerkbildung im Medium der Kulturzeitschrift“ untersuchten Kulturzeitschriften

Im Hinblick auf die Akteursnetzwerke folgt die Datenerhebung dabei zunächst einmal einer scheinbar offensichtlichen Herangehensweise. Die Forscher*innen notieren, welche Beiträger*in auf welcher Seite einer Kulturzeitschrift genannt wird. So lässt sich ein simples Netzwerk erstellen, das aufzeigt, wer wie viele Beiträge zu einer Zeitschrift in deren Erscheinungsverlauf beigesteuert hat. In der Forschung zu spanischsprachigen Kulturzeitschriften wurde aber bislang zumeist darauf verzichtet, mehr als *eine* der jeweils im nationalen Kontext kanonischen Kulturzeitschriften zu untersuchen. Dies hatte auch pragmatische Gründe, da die Forscher*innen bislang Bibliotheken und Archive aufsuchen mussten, die häufig nicht über breit ausdifferenzierte Sammlungen verfügen. Es ist aber die Verknüpfung der erhobenen Daten über mehrere Zeitschriften hinweg, die die Segmentierung der Archive überwindet, vor allem dann, wenn die Beiträge der Autor*innen zu mehreren Zeitschriften zeit- und kontinentübergreifend zusammengefasst werden. Dann zeigt sich der transkontinentale Kulturtransfer, der im Zentrum des Forschungsprojektes steht, und der auch schon jetzt tentativ über die Visualisierung eines kleineren Ausschnittes unseres Korpus anschaulich gemacht werden kann (Abb. 4). Wie in anderen Projekten der „Digital Humanities“ auch, ist für unseren Ansatz also die Verschiebung des Fokusses von der Analyse eines Einzelobjekts hin zu Vergleichsstrukturen bzw. Muster in einem großflächigeren, nicht mehr ausschließlich hermeneutisch lesbaren Korpus charakteristisch.

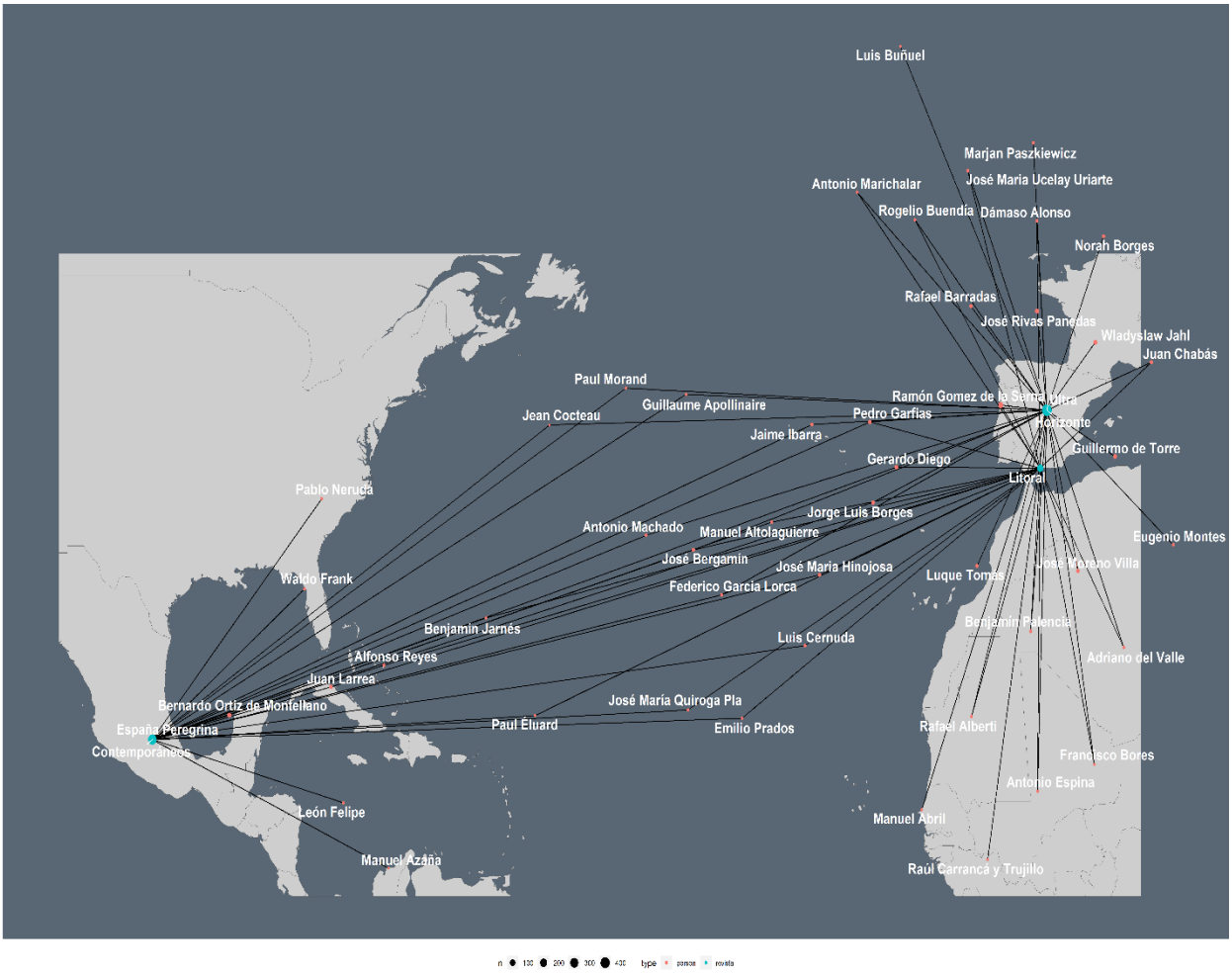

*Abb. 4: Visualisierung aller jener Beiträger*innen, die mehr als einen Beitrag zu mindestens zwei von fünf Kulturzeitschriften geleistet haben (Contemporáneos, España Peregrina – beide Mexico City; Horizonte, Ultra – beide Madrid; Litoral – Málaga)*

Unsere auf den transterritorialen und überstaatlichen Kulturtransfer abzielende Fragestellung rückt dabei nicht nur die relative Globalität der spanischsprachigen Moderne wieder in den Vordergrund, die z.B. in der englischsprachigen Forschung zu Kulturzeitschriften der Moderne weitgehend ausgeblendet wurde, sie wird auch ein vertieftes Verständnis eines der spezifisches Kennzeichen der Modernisierungsdynamik dieses Kulturraums ermöglichen, nämlich die wesentlich moderater ausfallende Disruptivität der literarischen Modernisierung. In den rhetorischen Selbstinszenierungen der Avantgarden in Europa dominierten bekanntlich Gesten des Bruches – der italienische Futurismus mit seinen permanenten Appellen zum radikalen Abriss aller „passatistischen" Kulturtraditionen einschließlich des literarischen Symbolismus ist dafür das schlagendste Beispiel. Im Verhältnis dazu scheint die explizite Anerkennung der Modernisierungsleistung der *modernistas* in den Zeitschriften der spanischsprachigen Avantgarden sehr viel dominanter als gezielte Invektiven gegen das vorangegangene literarische Paradigma, die eher vereinzelt zu finden sind. Soweit eine unserer zunächst hermeneutisch gewonnenen Vorannahmen, die wir durch unseren qualitative und quantitative Verfahren kombinierenden *mixed-method*-Ansatz neu und anders verifizieren wollen. Die bisherige Rede von der Datenerhebung verkürzte dabei allerdings das epistemische Problemfeld, das wir in diesem Beitrag fokussieren wollen und das den Visualisierungen, die wir in der Endphase des Projekts für das Gesamtkorpus vornehmen und analysieren werden, jeweils vorausliegt. Es ist gerade dieser, wie schon erwähnt extrem zeitaufwändige Prozess, der händisch und nur scheinbar quasimechanisch abläuft, dem wir hier Sichtbarkeit verschaffen wollen. Schon ein

erster genauerer Blick auf die Erhebung der Daten in der Forschungsumgebung zeigt bereits, wie problematisch der Begriff „Datum“ als „Gegebenes“ ist: Zwar liefert das IAI bereits Metadaten zu Autoren, Erscheinungsort oder verwendeten Genres, aber anonyme Beiträge, Beiträge in kollektiver Autorschaft oder Pseudonyme, wie sie in den Zeitschriften aus der Phase des *modernismo*, die oft unter den Bedingungen der Pressezensur erschienen, eher die Regel als die Ausnahme darstellten,[6] stehen im Gegensatz zu diesen scheinbar klaren Klassifikationen. Und auch die Erfassung der Grunddaten der Zeitschriften erweist sich in einigen Fällen als problematisch und eine alles andere als triviale Aufgabe. Einige Kulturzeitschriften wie „Alfar“ oder „Creación“ waren über ihre Funktion als Behältnisse für Texte unterschiedlicher kultureller Provenienz hinaus selbst mobile Artefakte, die aus unterschiedlichen Gründen von ihren Herausgebern an verschiedenen Orten herausgegeben wurden, so dass sich schon die basale Frage ergibt, ob sie als eine oder mehrere Zeitschriften gewertet werden müssen. So etwa die Zeitschrift *Alfar,* die von Julio C. Casal geleitet wurde, einem in Montevideo geborenen Sohn eines spanisch-uruguayischen Ehepaars, dessen diplomatische Karriere ihn zunächst nach Spanien führte, wo er die Zeitschrift im Galizischen *A Coruña* als Fortsetzung des *Boletín of the Casa América-Galicia* unter dem neuen Titel *Alfar* von Heft 33 (Oktober 1923) bis Heft 60 (1926) herausgab, bevor er dann das Projekt nach einer mehrjährigen Unterbrechung in Uruguay von 1929 bis 1954 fortsetzte. In Abwesenheit von Julio L. Casal edierte in A Coruña dann ein ehemaliger Mitarbeiter drei weitere Hefte, so dass es zwei konkurrierende Filiationslinien gibt. Diese Zeitschrift, die nicht nur für die Repräsentanz der nicht-kastilischen Literaturen der Iberischen Halbinsel seiner Zeit große Bedeutung hatte, sondern auch für den transatlantischen Kulturaustausch, ist in der konkreten Praxis der Forschung entsprechend unterschiedlich behandelt worden. Die bisher einzige in Spanien erschienene ausführliche Monographie behandelt sie als Zeitschrift ausschließlich im nationalen Kontext und verzichtet deshalb wie selbstverständlich auf die Nachgeschichte der Zeitschrift in Uruguay, [7] während die zentrale Plattform zur Präsentation digitaler Zeitschriften in Uruguay die Zeitschrift ebenso selbstverständlich ausschließlich Digitalisate aus der Phase in Uruguay bereitstellt.[8] Erklärt sich die Mobilität der Zeitschrift in diesem Fall allein aus den biographischen Umständen des Herausgebers, so ist die kulturübergreifende Mobilität im Fall von Huidobros zunächst mehrsprachig publizierter Zeitschrift *Creación / Création* auch inhaltlich programmatisch. Der chilenische Avantgardeautor ging in seiner Vision einer utopischen Gemeinschaft der Kunstschaffenden offenbar wie selbstverständlich davon aus, dass sprachliche und nationalstaatliche Grenzen für diese Internationale der Künste, die er in seinem Forum vernetzen wollte, keine Rolle mehr spielen. Bei der bibliothekarischen Klassifizierung einer solchen Zeitschrift muss dieser kulturüberschreitende Charakter entweder bewusst ignoriert werden, oder er wird durch normative pragmatische Zwänge notwendig diffundiert.

So trennt die *Bibliothèque Nationale de France* etwa die erste, mehrsprachig verfasste und in Spanien erschiene Nummer der Zeitschrift von den weiteren französischen Ausgaben und beschreibt sie als eine exotische, nicht näher erfasste abgeleitete Sekundärzeitschrift (Abb. 5 und 6). Das Iberoamerikanische Institut wiederum, das für unser Forschungsprojekt die Zeitschrift

[6] Deshalb existieren gerade für die Phase des *Modernismo* in den unterschiedlichen nationalen Kontexten des spanischsprachigen Kulturraums auch schon Hilfsmittel zur Auflösung der gängigsten Schriftstellerpseudonyme, die wir im Projekt natürlich nutzen.

[7] vgl. Molina, César Antonio: La revista alfar y la prensa literaria de su época (1920-1930), La Coruña: Ediciones Nos 1984.

[8] Anáforas: Publicaciones Periódicas del Uruguay. Revistas: Alfar. http://anaforas.fic.edu.uy/jspui/handle/123456789/5562 (Zugriffsdatum 20.11.2018).

digitalisiert hat, bietet seinerseits Metadaten dazu an, die ganz offensichtlich inkonsistent ausfallen, weil die Zeitschrift eindeutigen Klassifikationsprinzipien entgegensteht (vgl. Abb. 7).

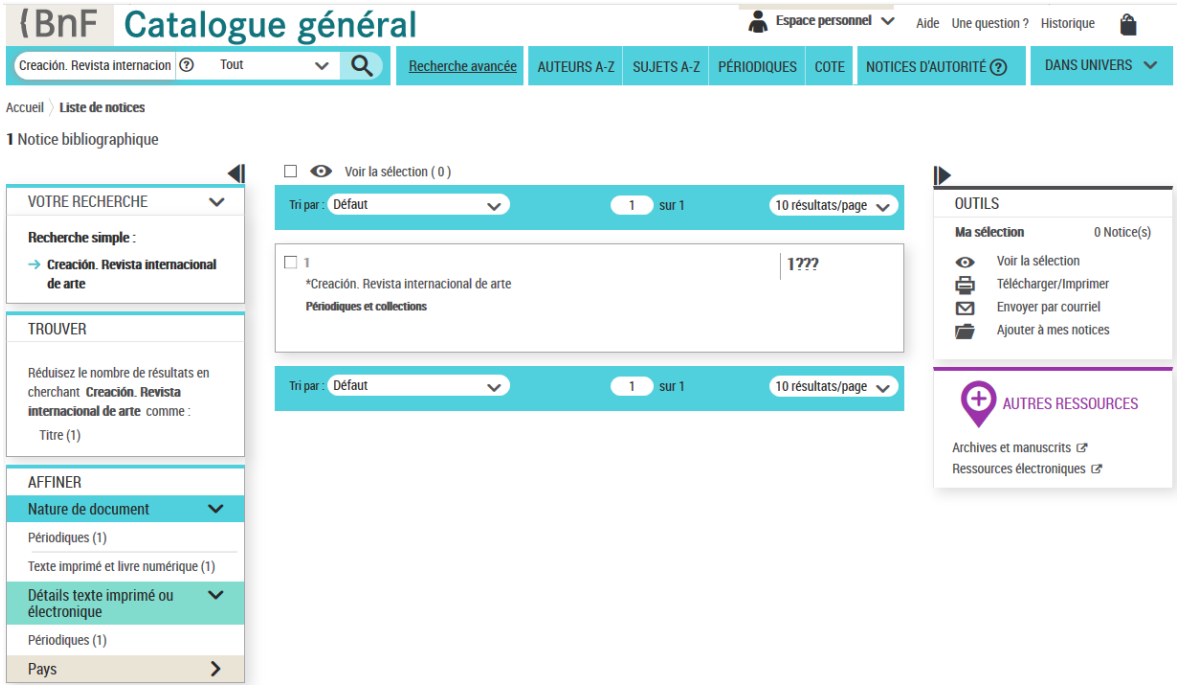

Abb. 5: Die Zeitschrift „Creación“, wie sie im Katalog der Bibliothèque Nationale de France (BnF) verzeichnet wird

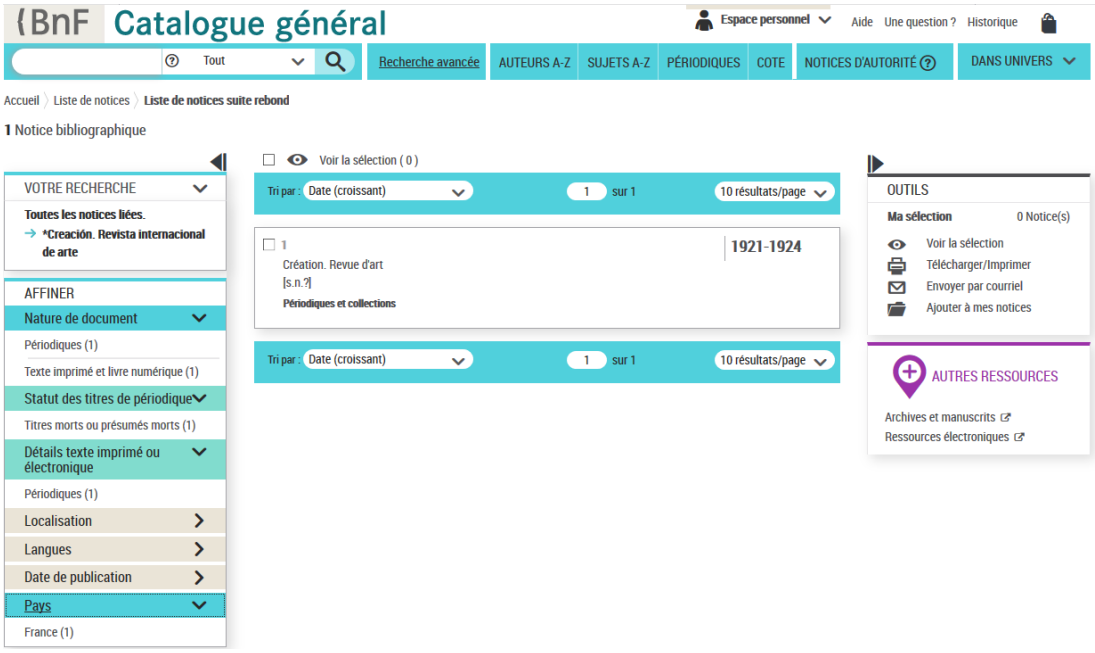

Abb. 6: Die Zeitschrift „Création“, wie sie im Katalog der Bibliothèque Nationale de France (BnF) verzeichnet wird

```
−<mets:mets OBJID="" xsi:schemaLocation="http://www.loc.gov/mods/v3 http://www.loc.gov/standards/mods/v3/mods-3-5.xsd
 v2-0.xsd http://www.loc.gov/standards/mix/ http://www.loc.gov/standards/mix/mix.xsd">
 −<mets:metsHdr CREATEDATE="2018-02-06T14:34:29Z">
  −<mets:agent OTHERTYPE="SOFTWARE" ROLE="CREATOR" TYPE="OTHER">
   −<mets:name>
      Goobi - ugh-3.0-ugh-2.0.0-29-g3b6efe1 - 21−December−2016
    </mets:name>
    <mets:note>Goobi</mets:note>
   </mets:agent>
  </mets:metsHdr>
 −<mets:dmdSec ID="DMDLOG_0000">
  −<mets:mdWrap MDTYPE="MODS">
   −<mets:xmlData>
    −<mods:mods>
     −<mods:originInfo>
      −<mods:place>
         <mods:placeTerm type="text">Madrid, Paris</mods:placeTerm>
       </mods:place>
       <mods:dateOther>Número 1 (abril 1921)-février 1924</mods:dateOther>
      </mods:originInfo>
      <mods:classification authority="ivdcc">Varia</mods:classification>
     −<mods:recordInfo>
       <mods:recordIdentifier source="gbv-ppn">1012316823</mods:recordIdentifier>
      </mods:recordInfo>
     −<mods:identifier type="purl">
       http://resolver.iai.spk-berlin.de/IAI0000695800000000
      </mods:identifier>
      <mods:identifier type="zdb">2918814-3</mods:identifier>
     −<mods:titleInfo>
       <mods:title>Creación</mods:title>
      </mods:titleInfo>
     −<mods:titleInfo type="alternative">
       <mods:title>Création <span.></mods:title>
      </mods:titleInfo>
     −<mods:subject authority="gnd">
       <mods:geographic>Spanien</mods:geographic>
      </mods:subject>
     −<mods:subject authority="gnd">
       <mods:geographic>Frankreich</mods:geographic>
      </mods:subject>
     −<mods:subject authority="gnd">
       <mods:topic>Kunst</mods:topic>
```


Abb. 7: Die Zeitschrift „Creación“, wie sie in den vom Ibero-Amerikanischen Institut (IAI) erhobenen Metadaten verzeichnet wird

Schon auf dieser ersten Ebene der Objekterfassung wird also klar, dass die Mehrdimensionalität des Objektes die Konsistenz des vorgegebenen und auf Permanenz ausgerichteten Klassifikationssystems unterminiert. Die dem Objekt geschuldete Ambiguität der geographischen Zuordnung wird unterschiedlich gelöst, wenn einerseits die Klassifikationskategorie „place“ durch den Doppeleintrag „Madrid – Paris“ ambivalent gehandhabt wird, oder wenn an anderer Stelle das „subject“ geographisch zweifach codiert wird. Dass dabei die Sprachlichkeit der Kategorienbildung (Englisch) mit der Kategorisierungssprache (Deutsch) kollidiert, ist allerdings nicht mehr dem Objekt geschuldet, sondern erklärt sich aus den Pragmatiken der Katalogisierung am IAI, wobei man sich schon die Frage stellen darf, wie sinnvoll es ist, einem internationalen Nutzerkreis, bei dem man sowohl spanische als auch englische Sprachkenntnisse voraussetzen kann, noch deutschsprachige Metadaten anzubieten.

Deutlich wird die epistemische Problematik der Kategorisierungen aber auch, wenn wir die Ebene der bibliothekarisch vorgegebenen Metadaten verlassen und jene Daten näher in den Blick nehmen, die von den Forscher*innen eigenständig zu erfassen waren, etwa die der Autor*innen der Beiträge. Scheinbar eindeutige „Daten“ wie Geschlecht oder Nationalität werden fragwürdig in einer Welt des 21. Jahrhunderts, in der jenseits der Binarität weiblich/männlich auch noch weitere Klassifikationen wie „Intersexualität“ Anwendung finden, oder wenn die Autor*innen migriert sind, wie es zahlreiche Kulturschaffende während des spanischen

Bürgerkriegs taten. In einer Welt der Daten, die auf zähl- und beobachtbare Einheiten fokussiert, werden Kategorien wie „männlich“, „weiblich“ oder „Staatsangehörigkeit“ als selbstverständlich vorausgesetzt und als ‚natürliche Unterschiede‘ gesehen, obwohl sie doch sozial konstruiert sind.[9] Polyvalente Terme und uneindeutige Begrifflichkeiten haben in solchen Klassifikationssystemen keinen Platz, oder höchstens in der Sammelkategorie „Andere“. Bereits an diesen Beispielen wird deutlich, dass ein positivistischer Zugang zur Wirklichkeit eine ‚neutrale Beobachtungssprache‘ postuliert und eine Doktrin von ‚reinen Daten‘ vorträgt, die unabhängig von Theorien existierten. Die Arbeit in der Forschungsumgebung hingegen verdeutlicht, dass *jede* Beobachtung theoriegebunden ist und daher keine ‚neutrale Beurteilung‘ ermöglicht wird. Vielmehr ist jede Beobachtung interessegeleitet, jedes Wissen ist sozial konstruiert.

Die gesellschaftliche Vereinbarung darüber, was als „Faktum“ oder „Datum“ angesehen wird, lässt sich am Beispiel der Genres besonders evident machen. Sicher kann man übergreifende Kategorien wie „Image“, „Review“, „Non-fictional prose“, „fictional prose“ o.ä. finden, denen jeder einzelne Beitrag zu einer Zeitschrift zugeordnet werden kann. Diese Kategorisierungen stehen aber häufig den von den Zeitschriften selbst verwendeten Klassifikationen entgegen, wie sie sich typischerweise in den Indizes finden, und die zusammengenommen eine inkommensurable Vielfalt von Genreklassifikationen aufbieten. Im Anschluss an Johanna Drucker[10] ist es daher sinnvoll, den Begriff „datum“ zu verwerfen und stattdessen „captum“ zu verwenden, etwas ‚Genommenes‘, das dem theoretischen Zugang entstammt, von dem aus das Untersuchungsmaterial in den Blick genommen wird. Dies eröffnet die Möglichkeit, Genres beispielsweise als einander überlappende Kategorien zu begreifen, deren genauere Bestimmung sich erst aus der Nachbarschaft mit anderen Genres im Kontext einer Zeitschrift ergibt, und die zusammengenommen das Profil einer Zeitschrift ausmachen (Abb. 8).

[9] Dazu ausführlich Westbrook, Laurel / Saperstein, Aliya: Rethinking the Measurement of Sex and Gender in Social Surveys. In: GENDER & SOCIETY 29,4 (2015), S. 534–560.

[10] Drucker, Johanna: Humanities Approaches to Graphical Display. In: DHQ 5,1 (2011). http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html (Zugriffsdatum: 20.11.2018).

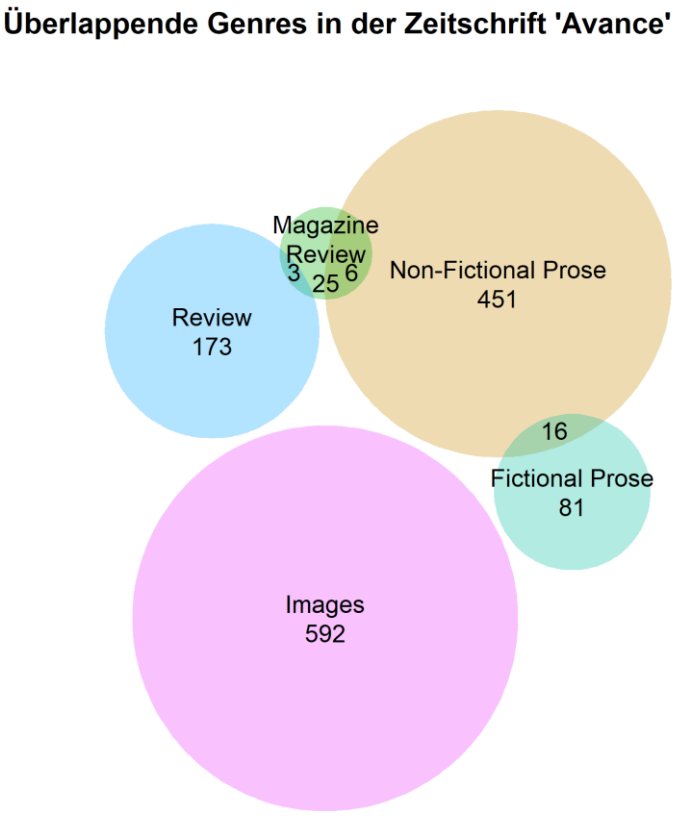

Abb. 8: Visualisierung der in der Kulturzeitschrift „Avance“ verwendeten Genres

Insgesamt ist es nicht überraschend, dass die am DFG-Projekt „Literarische Modernisierungsprozesse und transnationale Netzwerkbildung im Medium der Kulturzeitschrift“ beteiligten Forscherinnen ihr Genreklassifikationssystem permanent verändert und verfeinert haben. So gibt es insgesamt drei Spalten, in denen die verwendeten Genres beschrieben werden: In der Spalte „Genre Type: comparable“ werden Abstraktionen verwendet, die dem zeitschriftenübergreifenden Vergleich dienen, etwa „non-fictional prose“, „image“, oder „review“. In der Spalte „Genre Type: exact“ wird eine genauere Klassifikation verwendet; der generalisierende Begriff „non-fictional prose“ wird hier konkretisiert, etwa durch „Note“. In der dritten Spalte „Genre Type: specific“ schließlich wird die in der zweiten Spalte verzeichnete Begrifflichkeit spezifiziert; aus der „Note“ kann hier z.B. eine „programmatische“, „literarische“ oder „selbstreflexive Note“ werden. Das Wuchern dieser Taxonomien ist dabei unvermeidlich, ebenso wie die Diskussion problematischer Fälle, wie etwa die multiple Autorschaft eines Textes oder die Kombination verschiedener Genres in einem Textkonstrukt. An solchen Beispielen wird deutlich, dass Datennormalisierung, d.h. der Versuch, übergreifend gültige Klassifikationen zu finden, die keine Redundanzen mehr erhalten, zu einer Zwangsjacke wird. Gerade Literaturwissenschaftler*innen wird dies überdeutlich, sind sie doch Expert*innen für Polysemie und das

semantische Gleiten von Begriffen in ihrem Kontext. Und schließlich ist auch noch zu beachten, dass die Datenstruktur den Erfordernissen digitaler Auswertungsmöglichkeiten gehorchen muss, damit die Daten effektiv verwendet werden können (Abb. 9).

Type: comparable	Type: exact	Type: specific		Colour key
Image	Drawing	Charcoal Drawing		Needs more clarification = educated guess
Non-Fictional Prose	Comment	Editorial mission statement		Check further nationality
Review	Culture Review			No information
Fictional Prose	Short Story			Contributor name given but unreadable
Image	Drawing			
Lyricism	Poem			
Image	Drawing	Charcoal Drawing		
Non-Fictional Prose	Comment			
Lyricism	Prose			
Fictional Prose	Short Story			
Image	Drawing	Charcoal Drawing		
Image	Drawing	Charcoal Drawing		
Lyricism	Poem			
Lyricism	Poem			
Non-Fictional Prose	News tidbits			
Image	Drawing			
Non-Fictional Prose	Comment			
Image	Drawing			
Lyricism	Poem			
Fictional Prose	Book fragment			
Image	Drawing			
Non-Fictional Prose	Chronicle			
Lyricism	Poem			
Image	Drawing			
Review	Book review			
Image	Drawing			
Non-Fictional Prose	Comment			
Non-Fictional Prose	Letter			
Lyricism	Prose			
Non-Fictional Prose	Comment			

Abb. 9: Beispiel aus dem datasheet der Zeitschrift „Arte Joven" (mit Anmerkungen der Forscherin)

An dieser zunehmenden Ausdifferenzierung eines Klassifikationssystems wird deutlich, was bereits Ludwik Fleck in seinem klassischen Werk „Die Entstehung und Entwicklung einer wissenschaftlichen Tatsache"[11] ausgeführt hat: Die Wahrnehmung der Welt wird habitualisiert, d.h. auf eine ganz bestimmte Art und Weise zur Gewohnheit. Die Gewohnheit ist auch dafür zuständig, dass vergessen wird, dass ein Objekt qua Konvention und Erlerntem zur wissenschaftlichen Tatsache erklärt wurde, etwa wenn wiederkehrende Muster (*patterns*) erst sichtbar werden, wenn tausende von Zeitschriftenseiten untersucht wurden. Erworbene Wahrnehmungsgewohnheiten ermöglichen jene scheinbar wertneutrale, desinteressierte Beobachtung, die die Grundlage der Wissenschaft darstellt. Der Erwerb dieser Fähigkeiten ist den Forscher*innen zwar bewusst, wird aber als stillschweigendes Wissen hingenommen. Wie ein/e Wissenschaftler*in sehen zu lernen ist eine Frage von akkumulierter Erfahrung. In der virtuellen Forschungsumgebung mit ihren thesaurierenden Datenfeldern akkumuliert das Wissenschaftler-Kollektiv, nicht der Einzelne. Damit entwickeln und erlernen Kollektive von Wissenschaftlern Wahrnehmungsweisen als Teil ihrer *déformation professionelle*. Das Beispiel der

[11] Fleck, Ludwik: Entstehung und Entwicklung einer wissenschaftlichen Tatsache. Frankfurt am Main: Suhrkamp 1980. Vgl. dazu auch Daston, Lorraine: On Scientific Observation. In: Isis 99 (2008), S. 97–110.

Revistas Culturales zeigt damit auf eindrückliche Art und Weise auf, wie sich die Wahrnehmung dieser Klassifikationen durch Forscher*innen verändert, wenn hunderte oder gar tausende Seiten dieser Digitalisate betrachtet wurden.

3. Das Gewicht der Infrastruktur: Zur Interaktion zwischen Forschung und Bibliothekswesen im Zuge der Digitalisierung der Kulturgüter

Tritt man einen Schritt zurück und betrachtet das Zusammenspiel zwischen Forscherindividuum, Forschungsumgebung und den von den Bibliotheken bereitgestellten Metadaten, dann wird sichtbar, dass es sich hier um einen großflächigen Austauschprozess handelt. Sind es wirklich die Forscher*innen, die sich für eine bestimmte Art von Daten und spezifische Methoden entscheiden, oder wird diese Entscheidung durch die vorhandenen Objekte und die dafür geeigneten Werkzeuge und Methoden bereits präformiert?[12] Bereits bevor einzelne Forscher*innen Daten erheben, ist ihre Wahrnehmung von Kategorien durch die Herkunft dieser Daten aus Indizes und Registern vorformatiert. Diese Beobachtung lässt sich leicht an der Begeisterung der sog. ‚digital humanities' für Sach-, Themen- und Personenregister und Ortsverzeichnisse ablesen, die mit Techniken wie *topic modeling* oder *named entity recognition* korrespondieren. Oder auch die Vorliebe der ‚digital humanities' für Visualisierungen wie etwa Karten und Chronologien (Zeitstrahl). Alle diese Beispiele bezeichnen Ordnungs- und Kategorisierungssysteme, wie sie von jeher in historischen Büchern enthalten und damit der Erkenntnis vorgängig sind.

Es ist aber nicht nur die Welt der Bücher und die Ausbildung der Forscher*innen im universitären System, die die Wahrnehmung der Forscher*innen und die von ihnen durchgeführten epistemischen Prozesse formatieren. Vielmehr kommen hier über die – im Beispiel hier durch das IAI erhobenen – Metadaten der Bibliotheken mächtige Infrastrukturen ins Spiel. Verlage erstellen ihre Bücher als strukturierte Textdaten in Formaten wie SGML und speisen beschreibende Daten und Klappentexte in Vermarktungsmaschinerien wie das VLB ein. Bibliotheken akquirieren diese Metadaten im xml-Format und bereiten sie in ihren internen Datenbanken und den extern sichtbaren Bibliotheks-OPACs auf. Recherchen in diesen Instrumenten liefern präformierte Ergebnisse, die auf diesen Metadaten sowie auf den klassischen und historisch gewachsenen bibliothekarischen Standards basieren. Wenn man sich vor Augen führt, dass im Deutschen Reich, in der Weimarer Republik, in der DDR und der BRD zwischen 1899 und 1983 mit dem bibliothekarischen Regelwerk der Preußischen Instruktionen (PI) und zwischen 1976 und 2015 mit den Regeln für die alphabetische Katalogisierung (RAK) gearbeitet wurde und seit 2015 der Standard Resource Description & Access (RDA) verwendet wird, so muss man sich auch der Implikationen gewahr werden, die solche Regelwerke mit sich bringen. Um nur ein einziges eklatantes Beispiel zu nennen: Das international immer noch verwendete Regelwerk des Dewey Decimal Classification System (DDC) kennt zehn Kategorien für „Religion" (DDC 200). NEUN davon werden für christliche Religionen verwendet, während nur eine einzige (DDC 290) für ALLE nicht-christlichen Religionen und den Religionsvergleich verwendet werden kann.

[12] Gray, Jonathan / Gerlitz, Carolin / Bounegru, Liliana: Data infrastructure literacy. In: Big Data & Society (2018). html/pdf Format. DOI: 10.1177/2053951718786316 (Zugriffsdatum: 20.11.2018).

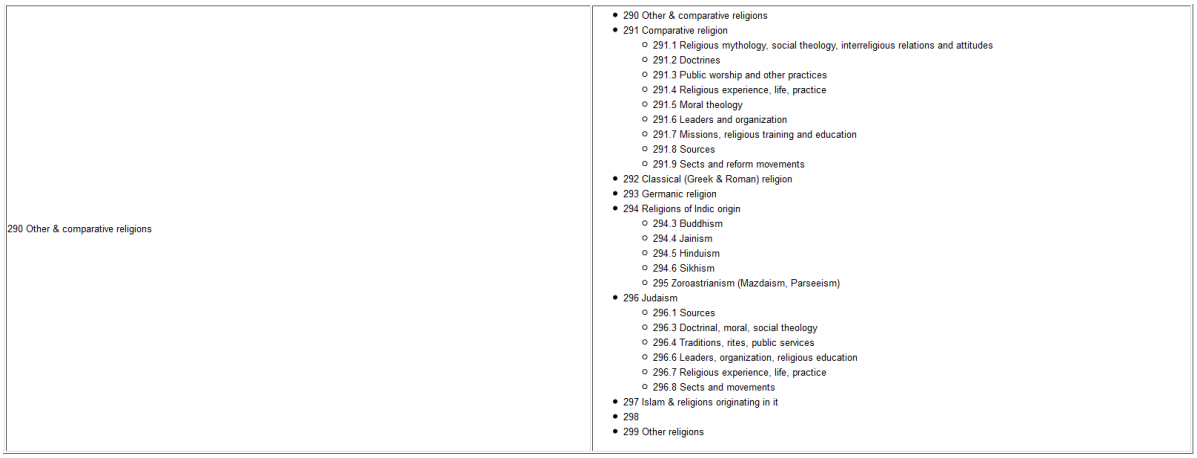

Abb. 10: Die Klassifikation DDC-290 „Other & comparative religions“

Bei einem Vergleich dieser Regelwerke würde sicherlich sofort die Inkommensurabilität dieser Klassifikationssysteme im Zeitverlauf deutlich werden. Hier wäre dann zu erörtern, inwiefern diese Klassifikationssysteme mit den zeitgenössischen Denkschemata ihrer Nutzer*innen korrespondieren. Schließlich müsste hinterfragt werden, in welchen Bereichen es große Übereinstimmungen zwischen den Klassifikationskategorien der Bibliotheken mit den von den Verlagen vergebenen Warengruppensystematiken gibt (etwa Bureau International des Containers BIC, Book Industry Subject and Category BISAC, dem xml-Standard Online Information eXchange ONIX, um nur diejenigen zu nennen, die am häufigsten verwendet werden). Das Beispiel Bibliotheken verdeutlicht, dass Klassifikationssysteme schwer zu analysieren sind – je besser sie funktionieren, und je besser sie von den Nutzern angenommen werden, desto weniger sichtbar sind sie. Und: Je umfassender und wirkmächtiger sie sind, umso weniger sind sie sichtbar.[13] Insgesamt werden solche Systeme daher als „gegeben“ (datum), unhinterfragbar und unveränderbar angesehen. Wie bereits Jean-François Lyotard 1982 prognostiziert hat, sind Datenbanken „die ‚Natur‘ des postmodernen Menschen“.[14]

Klassifikations- und Standardisierungssysteme formen Verbindungsstellen zwischen der sozialen Organisation einer Gesellschaft, der moralischen Ordnung, und den Systemen technischer Integration. Letztlich wird hier deutlich, dass im Zusammenspiel zwischen Forscherindividuum, Forschungsumgebung und von den Bibliotheken bereitgestellten Metadaten die Konstruktion von Wissen einen zyklischen Prozess darstellt: Die Überzeugungen über die Beschaffenheit der Welt fließen in das Design von Klassifikationssystemen ein, auf deren Grundlagen Daten erhoben und gesammelt werden. Die Suche in Bibliothekskatalogen und die Verwertung von Metadaten wiederum formt die Überzeugungen von Nutzer*innen und Wissenschaftler*innen über die Beschaffenheit der Welt.

Wie aber verhalten sich die in großflächigen Infrastrukturen vorhandenen Metadaten zu den in der virtuellen Forschungsumgebung erhobenen Daten? Technisch ist es möglich, die im DFG-Projekt erhobenen feingliedrigen Metadaten, in die das Fachwissen der auf das Material spezialisierten Forschercommunity eingeflossen ist, in das METS/MODS-Format der Bibliotheken zurückzuspielen. Unsere Erfahrung zeigt aber, dass die institutionelle Bereitschaft, von uns erhobenen Metadaten wieder aufzunehmen, gering ausgeprägt ist. Das hat zum einen

[13] Bowker, Geoffrey C. / Star, Susan Leigh: Sorting Things Out. Classification and Its Consequences. Cambridge MA / London: The MIT Press 1999.

[14] Lyotard, Jean François: Das postmoderne Wissen: Ein Bericht. 3. unveränderte Neuaufl. Wien: Passagen Verlag 1994, 151.

pragmatische Gründe, z.B. mangelnde personelle Ressourcen; zum anderen aber auch technische, wenn man etwa an die Notwendigkeit der Versionierung und der Bereitstellung wachsender Datenspeicher denkt. Zwar ist das METS/MODS-Format flexibel und offen für die Anreichung der Datensätze. Indes wandern die in unserem kleinen, wendigen Portal entwickelten flexiblen, materialadäquaten klassifikatorischen Antworten deshalb noch lange nicht in die bibliothekarischen Klassifikationssysteme ein. Vielmehr beweisen diese eine große Beharrungskraft und die Formationsmacht der Tradition, die den dynamischen Entwicklungen innerhalb der Forschung entgegenstehen. So bleibt festzuhalten, dass Dateninfrastrukturen auch weiterhin als normative Kraft wirken können.[15] Digitale Infrastrukturen prägen damit insgesamt den Denkraum, innerhalb dessen geforscht wird. Dies geschieht zum einen bei der Recherche und Auswahl der zu untersuchenden Materialien, zum anderen durch die Formatierung der Wahrnehmung der Forscher*innen. In unserem Beispiel wirken die Bibliotheks-Metadaten auf das Wissenschaftlerkollektiv zurück, da die METS/MODS-Dateien bereits Genreklassifikationen enthalten. Diese Etikettierungen beeinflussen die Text- bzw. Zeitschriftenauswahl durch die Forscher und ihr Verständnis der Charakteristika von Genres und sie verstetigen damit die Kanonbildung. Ein kritisches Bewusstsein dieser normativen Macht und die Bereitschaft, diese zu hinterfragen bzw. zu relativieren, ist aber auch mit forschungspolitischen Apellen zu einer stärkeren Orientierung an der Erbringung von Dienstleistungen für die Forschung, wie sie mit der Umstellung von Sondersammelgebieten zu Fachinformationsdiensten von der DFG auf den Weg gebracht wurden, allein noch nicht zu erreichen. Es bleibt also abschließend zu fragen, wie denn der Dialog zwischen Forschung und Bibliothekswesen in Zukunft auszusehen hätte, in dem die beschriebenen Friktionen produktiv gemacht werden können.

Literaturverzeichnis

Anáforas: Publicaciones Periódicas del Uruguay. Revistas: Alfar. http://anaforas.fic.edu.uy/jspui/handle/123456789/5562 (Zugriffsdatum 20.11.2018).

Bilbao, Francisco: Post-Dictum. In: Manuel Bilbao (Hg.): Obras completas de Francisco Bilbao, Vol. 1, Buenos Aires: Imprenta de Buenos Aires 1866, S. 285–304.

Bowker, Geoffrey C. / Star, Susan Leigh: Sorting Things Out. Classification and Its Consequences. Cambridge MA / London: The MIT Press 1999.

Daston, Lorraine: On Scientific Observation. In: Isis 99 (2008), S. 97–110.

Drucker, Johanna: Humanities Approaches to Graphical Display. In: DHQ 5,1 (2011). http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html (Zugriffsdatum: 20.11.2018).

Fleck, Ludwik: Entstehung und Entwicklung einer wissenschaftlichen Tatsache. Frankfurt am Main: Suhrkamp 1980.

[15] Gray et al. 2018, S. 3.

Gray, Jonathan / Gerlitz, Carolin / Bounegru, Liliana: Data infrastructure literacy. In: Big Data & Society (2018). html/pdf Format. DOI: 10.1177/2053951718786316 (Zugriffsdatum: 20.11.2018).

Lehmann, Jörg / Huber, Elisabeth / Stodulka, Thomas: H2020 Project K-PLEX: WP4 Report on Data, Knowledge Organisation and Epistemics, Research Report, Berlin: Freie Universtät Berlin 2018. HAL Id: <hal01761214>. https://hal.archives-ouvertes.fr/hal-01761214 (Zugriffsdatum: 20.11.2018).

Lyotard, Jean François: Das postmoderne Wissen: Ein Bericht. 3. unveränderte Neuaufl. Wien: Passagen Verlag 1994.

Molina, César Antonio: La revista alfar y la prensa literaria de su época (1920-1930), La Coruña: Ediciones Nos 1984.

Moretti, Franco: Conjectures on World Literature. In: New Left Review 1 (2000), S. 54–68.

Moretti, Franco: Operationalizing. Or, the Function of Measurement in Literary Theory. In: New Left Review 84 (2013), S. 103–119.

Trilcke, Peer / Fischer, Frank: Literaturwissenschaft als Hackathon. Zur Praxeologie der Digital Literary Studies und ihren epistemischen Dingen. In: Martin Huber / Sybille Krämer (Hg.): Wie Digitalität die Geisteswissenschaften verändert: Neue Forschungsgegenstände und Methoden, 2018 (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 3). text/html Format. DOI: 10.17175/sb003_003 (Zugriffsdatum: 20.11.2018).

Westbrook, Laurel / Saperstein, Aliya: Rethinking the Measurement of Sex and Gender in Social Surveys. In: GENDER & SOCIETY 29,4 (2015), S. 534–560.

Abbildungsverzeichnis

Abb. 1: Screenshot der virtuellen Forschungsumgebung Revistas Culturales 2.0 mitsamt den biblio-thekarischen Metadaten; gewähltes Beispiel: Die Kulturzeitschrift „Plus Ultra“. © Eigene Grafik, 2018.

Abb. 2: Screenshot der Arbeitsumgebung in der virtuellen Forschungsumgebung Revistas Culturales 2.0; eine einzelne Seite der Kulturzeitschrift „Plus Ultra“. © Eigene Grafik, 2018.

Abb. 3: Visualisierung der räumlichen Verteilung der fast 60 im DFG-Projekt „Literarische Modernisierungsprozesse und transnationale Netzwerkbildung im Medium der Kulturzeitschrift“ untersuchten Kulturzeitschriften. © Eigene Grafik, 2018.

Abb. 4: Visualisierung aller jener Beiträger*innen, die mehr als einen Beitrag zu mindestens zwei von fünf Kulturzeitschriften geleistet haben (Contemporáneos, España Peregrina – beide Mexico City; Horizonte, Ultra – beide Madrid; Litoral – Málaga). © Eigene Grafik, 2018.

Abb. 5: Die Zeitschrift „Creación“, wie sie im Katalog der Bibliothèque Nationale de France (BnF) verzeichnet wird. © Eigene Grafik, 2018.

Abb. 6: Die Zeitschrift „Création“, wie sie im Katalog der Bibliothèque Nationale de France (BnF) verzeichnet wird. © Eigene Grafik, 2018.

Abb. 7: Die Zeitschrift „Creación“, wie sie in den vom Ibero-Amerikanischen Institut (IAI) erhobenen Metadaten verzeichnet wird. © Eigene Grafik, 2018.

Abb. 8: Visualisierung der in der Kulturzeitschrift „Avance“ verwendeten Genres. © Eigene Grafik, 2018.

Abb. 9: Beispiel aus dem datasheet der Zeitschrift „Arte Joven“ (mit Anmerkungen der Forscherin). © Eigene Grafik, 2018.

Abb. 10: Die Klassifikation DDC-290 „Other & comparative religions“. © Eigene Grafik, 2018.

Konzepte digitaler (Re-)Präsentationen von Literatur zwischen Pluralisierung und Standardisierung

Julia Nantke (Wuppertal)

1. Einleitung

Die Digitalisierung der Literaturwissenschaft geht mit einer zunehmenden Verlagerung ihrer literarischen Untersuchungsgegenstände in den digitalen Raum einher. Diese ist zum einen pragmatisch durch die Notwendigkeit einer digitalen Verfügbarkeit für computergestützte Analysen motiviert und wird zum anderen affirmativ begründet durch die Möglichkeiten eines ebenso umfassenden wie ubiquitären Zugriffs auf Literatur jenseits räumlicher, materieller und institutioneller Grenzen, die im Umkehrschluss wiederum eine computergestützte Untersuchung geradezu einfordern.[1]
Indem jene literarischen Texte, die das maßgebliche Ziel (digitaler) literaturwissenschaftlicher Untersuchungen bilden, zunächst vorrangig in analoger Form vorliegen, kommen bei deren Überführung auf digitale Plattformen vielfältige Transformationsmechanismen zum Tragen, deren epistemisches Potential mit dem Postulat einer im Gegensatz zu analogen Repräsentationsformen realisierbaren Vollständigkeit[2] nicht hinreichend erfasst ist.[3] Dies liegt zum einen an der Vielfalt der materiellen Formate, welche die Grundlagen digitaler Infrastrukturen bilden. Zum anderen begründet sich dies durch die Spezifika digitaler Medialität, welche sich auf die (Re-)Präsentationsmöglichkeiten und -konzepte im Rahmen digitaler Plattformen auswirken.[4]

Unter Einbezug von Beispielen verschiedener Formate der Instanziierung von Literatur in digitalen Archiven und Editionen werden im Folgenden die „Resemantisierungen“[5] analoger Literatur in digitalen Infrastrukturen untersucht. Es soll gezeigt werden, dass sich diese aktuell in einem Spannungsfeld pluralisierender und standardisierender Tendenzen vollziehen, wobei sie als Produkte wie Ausgangsbedingungen literaturwissenschaftlicher Forschung deren Episteme in entscheidender Weise prägen. Die „typographischen Dispositive“[6] der Druckkultur werden dabei einerseits durch Strukturen abgelöst, die sich an den neuen Möglichkeiten digitaler (Re-)Präsentation orientieren, aber gleichermaßen durch die spezifischen Bedingungen digitaler Medialität determiniert sind. Andererseits interferieren diese digitalen Strukturen auf vielfältige Weise mit jenen der Drucke und Manuskripte, die jeweils als Repräsentanten der literarischen Texte digital instanziiert werden.
Die Verdoppelung der Untersuchungsperspektive auf die Präsentation ebenso wie die Repräsentation von Literatur trägt der Tatsache Rechnung, dass die betrachteten Infrastrukturen zum einen darauf zielen, Literatur im Internet sicht- und verfügbar zu machen, dabei aber zum

[1] vgl. Jockers, Matthew J.: Macroanalysis. Digital Methods & Literary History. Urbana u. a.: University of Illinois Press 2013, S. 7f.
[2] vgl. Plachta, Bodo: Editionswissenschaft. Stuttgart: Reclam 2006, S. 135.
[3] vgl. hierzu die Anmerkungen zu „remediation as reform“ in Bolter, Jay David / Grusin, Richard: Remediation. Understanding New Media. Cambridge / London: MIT Press 2000, S. 59–62.
[4] vgl. Manovich, Lev: The Language of New Media. London: MIT Press 2001, S. 46f.
[5] Jäger, Ludwig: Intermedialität – Intramedialität – Transkriptivität. In: Arnulf Deppermann / Angelika Linke (Hg.): Sprache intermedial. Stimme und Schrift, Bild und Ton. Berlin: De Gruyter 2010, S. 301-324, hier S. 304.
[6] vgl. Wehde, Susanne: Typographische Kultur. Eine zeichentheoretische und kulturgeschichtliche Studie zur Typographie und ihrer Entwicklung. Tübingen: De Gruyter 2000, Kap. 5.1.1.

anderen die präsentierten Gegenstände in spezifischer Form repräsentieren, was einem Herstellungsprozess in konzeptueller wie technischer Hinsicht entspricht.[7]
Hierbei werden neben medialen Aspekten gleichermaßen theoretisch-methodologische Perspektiven relevant. Deshalb werden im Folgenden sowohl die Instanziierungsformen einzelner Texte als auch deren Einbettung in textuelle Netzwerke genauer beleuchtet. In beiden schlagen sich – so die These – spezifische Konzepte von literarischer Textualität nieder. Indem die Formen digitaler (Re-)Präsentation von Literatur noch keine den analogen Dispositiven vergleichbar stabilen Formate ausgebildet haben, tragen die Darstellungs-, Selektions- und Verknüpfungsmechanismen entscheidend zur Ausprägung digitaler Paradigmen der (Re-)Präsentation und damit der Wahrnehmung von Literatur bei.
Da eine Untersuchung im Rahmen eines kurzen Beitrags nur schlaglichtartig ausfallen kann, besteht das Ziel weniger in einer umfassenden Sichtung, als vielmehr im Abstecken zentraler Untersuchungsfelder, an die sich eine detailliertere und umfänglichere Analyse der hier aufgeworfenen Phänomene erst anzuschließen hätte. Nach einer kurzen Reflexion zu den in diesem Beitrag anvisierten Untersuchungsgegenständen (2.) werden als zentrale Felder die Maßgaben der Textauswahl und -zusammenstellung (3.), Charakteristika der (Re-)Präsentation der Texte (4. und 5.) sowie die durch die Infrastrukturen nahegelegten Lektüremodi (6.) beleuchtet.[8]

2. Mediale und systematische Transgressionen

Die Untersuchungen im Rahmen dieses Beitrags bewegen sich im Grenzbereich digitaler Archive und Editionen, für die im Zuge der Digitalisierung beider Formate eine zunehmende Konvergenz in Ausrichtung, Präsentationsformen und Nutzungsperspektiven bereits einschlägig beschrieben worden ist.[9] Neben dieser Grenzverwischung in systematischer Hinsicht ist ein Ergebnis dieses Prozesses, dass konkrete ‚Gattungszuordnungen' von Seiten der Projekte selbst teilweise vermieden und im Titel durch Alternativen wie „Portal", „Projekt" oder die Ergänzung „online" bzw. „digital" ersetzt[10] oder die Transgressionen von den Projektverantwortlichen explizit benannt werden.[11] Edition und Archiv sind im Digitalen also keine distinkten Konzepte mehr, sondern deren Leitideen der umfänglichen, strukturierten Sammlung und

[7] vgl. hierzu Benthien, Claudia / Klein, Claudia: Praktiken des Übersetzens und Rahmens. Zur Einführung. In: Dies (Hg.): Übersetzen und Rahmen: Praktiken medialer Transformation. Paderborn: Fink 2017, S. 9–16, hier S. 19, welche Praktiken medialer Rahmung als Praktiken der „Aktualisierung, Re-Formulierung, Transformation oder Neukonstruktion von Rahmungen" in den Blick nehmen; Jäger 2010, S. 313f. sowie allgemein auch Latour, Bruno: Drawing Things Together. In: Michael E. Lynch / Steve Woolgar (Hg.): Representation in Scientific Practice. Cambridge: MIT Press 1990, S. 19–68.
[8] Aus Platzgründen weitgehend ausgeblendet werden in diesem Beitrag wirtschaftliche und rechtliche Aspekte der Digitalisierung kultureller Artefakte. Diese bilden für die adäquate Erfassung der Entwicklung digitaler Infrastrukturen zweifelsohne sehr relevante Faktoren, die in einer umfassenden Untersuchung unbedingt zu berücksichtigen wären.
[9] vgl. u. a. Sahle 2007; Lukas 2013, S. 38f. sowie kursorisch Nutt-Kofoth 2013, vgl. dagegen aber auch Robinson 2009, S. 48f.
[10] vgl. z. B. *Ernst Jandl Online* (http://jandl.onb.ac.at), *Handke online* (https://handkeonline.onb.ac.at), *Heinrich-Heine-Portal* (http://germazope.uni-trier.de:8080/Projekte/HHP/), *Parzival-Projekt* (http://www.parzival.unibe.ch/home.html). Der Umstand, dass v. a. im anglo-amerikanischen Raum der Trend zur Konvergenz durch Benennungen wie *Walt Whitman Archive*, *Rossetti Archive* uvm. noch gestärkt wird (vgl. hierzu Nutt-Kofoth 2017, S. 110), verweist auch darauf, dass die hier ausschließlich in Bezug auf deutschsprachige Projekte angestellten Beobachtungen in internationaler Perspektive zu erweitern wären.
[11] vgl. Wizisla 2000.

Pflege literarischer Nachlässe einerseits sowie von deren wissenschaftlicher Aufbereitung und Publikation andererseits bilden vielmehr Merkmale, die mit unterschiedlichen Schwerpunktsetzungen für digitale wissenschaftliche Infrastrukturen veranschlagt werden können, die literarischen Inhalten gewidmet sind.[12]
Im Rahmen dieses Beitrags wurden daher Projekte berücksichtigt, die an verschiedenen Punkten einer Skala zwischen Archiv und Edition verortbar sind. Hierbei bestehen unterschiedliche ‚Aggregatszustände' nicht nur im Hinblick auf die medialen Formate analoger Artefakte, die durch den Computer als ‚Intermedium' (re)präsentiert werden können: Neben den Ausgangsobjekten variieren auch die Zielmedien zwischen rein digitalen Ausgaben und verschiedenen Hybrid-Typen, an die sich unterschiedliche Nutzungsszenarien knüpfen.

Zudem befinden sich viele der im Folgenden betrachteten Projekte noch im Entwicklungsstadium. Während für Archive die kuratorische Arbeit ohnehin nicht im Sinne eines definitiven Abschlusses perspektiviert werden kann, unterliegen gedruckte Editionen normalerweise durch ihre Publikation zeitlich und systematisch einer ‚natürlichen Begrenzung'. Für digitale Publikationen gelten hingegen nicht nur gesteigerte räumliche Expansionsmöglichkeiten, sondern dies gilt ebenso in zeitlicher Perspektive. Daran knüpfen sich sowohl Vorteile, etwa im Hinblick auf die Einarbeitung von Korrekturen und Ergänzungen sowie die nachträgliche Aufnahme relevanter Forschungsbeiträge, als auch Nachteile bezüglich der für wissenschaftliche Zwecke notwendigen Stabilität der durch Archive und Editionen bereitgestellten Forschungsgrundlagen. Letztere gelten für die Benutzung der Infrastrukturen im Rahmen der wissenschaftlichen Analyse literarischer Inhalte ebenso wie für die hier vorgenommene Analyse ihrer Darstellungs- und Herstellungsformen selbst.[13] Gleichzeitig bilden die fortgesetzten Entwicklungsmöglichkeiten einen entscheidenden Aspekt in der Reflexion der Strukturen digitaler wissenschaftlicher Infrastrukturen, der sich auf Produktions- wie Benutzungsmodalitäten auswirkt.[14] Im Rahmen dieses Beitrags werden deshalb Projekte betrachtet, bei denen zumindest soviel Material online vorhanden ist, dass deren grundsätzliche Ausrichtung und die konzeptuellen Strukturen der Materialaufbereitung sichtbar werden. Dies geschieht unter dem Vorbehalt, dass die anhand der Projekte entwickelten Überlegungen großenteils auf Momentaufnahmen basieren und nachträgliche Änderungen immer auch Umperspektivierungen nach sich ziehen könnten. Wo in Vorworten der Projekte o. ä. weitere geplante Funktionen thematisiert werden, werden diese jeweils in die Überlegungen einbezogen.

3. Zentrierte und kontextualisierende Korpora

Einen zentralen Aspekt der Raumerweiterung im Zuge der Digitalisierung wissenschaftlicher Archive und Editionen bildet die Multiplikation von Perspektiven auf die (re)präsentierten Gegenstände in Form von unterschiedlichen Zusammenstellungen, und auf den Plattformen vorgesehenen Zugriffsmöglichkeiten. Bedingen materielle Speicher die Vorherrschaft *eines* organisierenden Prinzips und somit auch *einer* übergeordneten Perspektive auf das dargebotene

[12] So wird bspw. die Plattform *Handke online* als „virtuelle[s] Archiv" perspektivert, gleichzeitig aber die Nachverfolgung der „Entwicklung der jeweiligen Schreibprojekte [...] bis hin zur Buchpublikation" als Ziel ausgegeben (https://handkeonline.onb.ac.at). Umgekehrt enthält die *Faustedition* einen eigenen Reiter „Archiv", unter dem die Handschriften sowie die „zu Lebzeiten erschienen textkritisch relevanten Drucke zum ‚Faust'" versammelt sind (http://beta.faustedition.net).
[13] vgl. hierzu im Hinblick auf die *Faustedition* Gengnagel 2017.
[14] vgl. Sahle 2016, S. 29f.

Material, kann im Kontrast dazu die Multiplikation selbst geradezu als digitaler Standard bewertet werden,[15] der in unterschiedlicher Form realisiert wird.

Bereits die grundsätzliche Anlage der jeweiligen Textsammlung verweist hier auf konkrete Vermittlungskonzepte und antizipierte Nutzungsszenarien. Dabei lassen sich anhand zweier gegenläufigen Konzepte grundsätzliche Tendenzen digitaler (Re-)Präsentation ablesen: Zum einen wird die philologische Ausrichtung auf die singuläre Autorin durch die mediale Raumerweiterung noch gestärkt, zum anderen scheint sich ergänzend ein zweites (Re-)Präsentationskonzept zu etablieren, welches die Netzwerkhaftigkeit digitaler Medialität im Sinne einer stärker thematisch orientierten Sammlung nutzt.

So ist bspw. die *Faustedition* in ihrer Anlage am klassisch-literaturwissenschaftlichen Parameter der Werk-Edition ausgerichtet. Der dabei vertretene textgenetische Ansatz kann im Rahmen der digitalen Aufbereitung als archivarische Akkumulation aller zum *Faust* verfügbaren Materialien realisiert werden, die hier allerdings durch die traditionell eng mit dem Werkkonzept verknüpfte Kategorie der Autorschaft begrenzt werden.[16] Auf der Plattform *Briefe und Texte aus dem intellektuellen Berlin um 1800* hingegen wird die präsentierte Auswahl nicht personell, sondern räumlich begrenzt: Sie „versammelt Texte verschiedener Autoren und unterschiedlicher Gattungen, die Eines gemeinsam haben: Sie beleuchten auf einprägsame Weise das intellektuelle Leben im Berlin des späten 18. und frühen 19. Jahrhunderts.“[17] Auch diese Ausrichtung lässt sich im Sinne der Annäherung von Edition und Archiv interpretieren, indem die Zusammenhangstiftung über den räumlich-thematischen Bezugspunkt als Alternative zum Autor-bezogenen Sammlungskonzept bereits für analoge Archive ein etabliertes Format darstellt, wohingegen analoge literarische Editionen sich nahezu ausschließlich am Primat der Autorschaft orientieren.[18]
Sowohl der textgenetische als auch der räumlich-thematische Ansatz implizieren eine Multiplikation von Materialien und Zugriffsweisen und tragen zu einer Annäherung von Edition und Archiv bei, welche allerdings auf verschiedenen Ebenen erfolgt und entsprechend mit unterschiedlichen Schwerpunktsetzungen verbunden ist. So transportiert die *Faustedition* eine auf den Nukleus Autor/Werk zentrierte Perspektive, die Goethes *Faust* in allen seinen entstehungsgeschichtlichen Facetten auffächert, und Ähnliches gilt für die ebenfalls textgenetisch angelegte Edition der *Notizbücher Fontanes* sowie die Plattform *Handke online*. Die *Briefe und Texte aus dem intellektuellen Berlin* sowie die „digitale Quellenedition“ der Avantgardezeitschrift *Der Sturm* realisieren hingegen jeweils eine historisch-kontextualisierende Perspektive, die ausgehend vom gemeinsamen thematischen Schwerpunkt den Blick auf die Vielfalt der an diesen geknüpften geistigen und künstlerischen Aktivitäten öffnet. Verschiedene Einstiege gruppieren hierbei das Material nach Autoren, Gattungen sowie Themen (*Briefe und Texte*) bzw. Orten (*Sturm*). Ebenfalls ein historisch-kontextualisierender Ansatz liegt der Textsammlung des *Deutschen Textarchivs* (DTA) zugrunde. Ziel ist es, auf der Basis „einer von Akademiemitgliedern erstellten und ausführlich kommentierten, umfangreichen Bibliographie“ eine

[15] So nennt Sahle 2016, S. 27 als einen Aspekt des von ihm postulierten „digital paradigm“, welches eine digitale von einer lediglich digitalisierten Edition unterscheidet „its representation of a potentially large number of documents in a potentially limitless number of different views“.
[16] vgl. hierzu auch Nantke 2018.
[17] https://www.berliner-intellektuelle.eu
[18] Bender 2016, S. 61 erklärt selbst noch im Hinblick auf digitale Editionen, dass der „inhaltlich-thematische Zusammenhang einer Edition [...] zum einen *offensichtlich* darin [besteht], dass alle ihre textuellen Bestandteile in Bezug zu einem bestimmten Autor und seinem Werk sowie einem entsprechenden historischen und narrativen Hintergrund stehen“. (Hvh. J. N.) Dies zeigt die anhaltende Wirkmächtigkeit jenes Paradigmas, welches allerdings durch die angeführten Gegenbeispiele konterkariert wird.

„Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache“ zu schaffen.[19] Literarische Texte finden sich hier entsprechend ebenso verortet wie Gebrauchsliteratur und wissenschaftliche Texte, wobei insbesondere solche Texte gesammelt werden, „die überregional wirksam waren“.[20] Die „Zeitleiste[n]“-Ansicht als zentraler Einstieg neben der alphabetischen Auflistung und der Suche vermittelt der Rezipientin dabei einen (durch die getroffene Textauswahl determinierten) Eindruck der Publikationsgeschichte des 19. Jahrhunderts.[21]

Die Beispiele zeigen, dass insbesondere die kontextualisierende Ausrichtung digitaler Infrastrukturen alternativer Konzepte der Zusammenhangstiftung jenseits von Autor und Werk bedarf. Die zitierte Zielsetzung des DTA verweist dabei darauf, dass diese Notwendigkeit wiederum eine Standardisierung auf Ebene der Textauswahl befördern kann: Die Digitalisierung ermöglicht zwar potentiell das Verfügbarmachen sämtlicher analog vorliegender Schriftartefakte, in der Praxis bedingen aber auch hierbei umsetzungstechnische Faktoren sowie systematische Erwägungen eine Auswahl, die entweder auf der Basis eines klar umrissenen thematischen oder räumlichen Fokus erfolgen kann oder – wie beim DTA – vorrangig zugunsten bereits einschlägig kanonisierter Texte ausfällt, die auf diese Weise in ihrer Bedeutung als ‚Standardtexte‘ literaturwissenschaftlicher Forschung stabilisiert werden.[22]

Die durch die digitalen Plattformen vermittelten unterschiedlichen Perspektiven auf die (re)präsentierten Objekte erscheinen weniger durch die Spezifika der jeweiligen Materialien, denn durch deren literaturhistorischen Status bzw. ein konkretes theoretisch-methodologisches Konzept motiviert. Davon zeugt nicht nur die Einbindung der Texte in ein Netzwerk aus kanonischen Kotexten im DTA. Ebenso bildet die traditionell enge Verknüpfung von Autorschaft und Werk den zentralen Orientierungspunkt für *Handke online*, welches sich zwar als „virtuelles Archiv“ versteht, in dem aber „nur jene Materialien [erfasst werden], die für seine als Buch veröffentlichten Werke relevant sind“.[23] Paradigmatisch zeigt sich dies ebenfalls im Vergleich der minutiösen Darbietung der Entstehungsgeschichte des in autorisierter Endfassung vorliegenden *Faust*, der die verhältnismäßig reduzierte Präsentation von noch unedierten Texten in den *Briefen und Texten aus dem intellektuellen Berlin* gegenübersteht. Auch hierbei scheint der Kanonisierungsgrad die Form und den Umfang der digitalen Aufbereitung zu beeinflussen. Insgesamt kann dies aktuell auch als generelle Tendenz der digitalen Textaufbereitung veranschlagt werden. So konzentrieren sich nicht nur die großen Textsammlungen des DTA und der *TextGrid*-Bibliothek explizit auf den Kanon,[24] sondern dies gilt auch für die digitalen (Re-)Präsentationen einzelner (Gesamt-)Werke. Diese beziehen sich im deutschsprachigen Raum momentan nahezu ausschließlich auf bereits kanonisierte männliche Autoren.[25]

[19] http://www.deutschestextarchiv.de.
[20] http://www.deutschestextarchiv.de/doku/textauswahl.
[21] vgl. http://www.deutschestextarchiv.de/list/timeline.
[22] Die aktuell im DTA verfügbare Textauswahl wirft allerdings die Frage auf, inwieweit das postulierte Ziel des Referenzkorpus tatsächlich realisiert wird. Der Umstand, dass sich das Korpus zu nicht geringen Teilen aus den Forschungsergebnissen fremder Projekte speist, trägt hier teilweise zu einer Erweiterung des nach spezifischen Kriterien selegierten Kanons bei.
[23] https://handkeonline.onb.ac.at/node/2594
[24] Ähnlich wie auf der DTA-Seite wird auch auf der Webseite von *TextGrid* darauf verwiesen, dass die digitale Bibliothek „nahezu alle wichtigen kanonisierten Texte und zahlreiche weitere literaturhistorisch relevante Texte enthält, deren urheberrechtliche Schutzfrist abgelaufen ist“ (https://textgrid.de/digitale-bibliothek).
[25] vgl. hierzu auch Lukas 2013, S. 44.

4. Visuelle und strukturelle Textvervielfachungen

Hinsichtlich der untersuchten Infrastrukturen lassen sich zwei verschiedene Formen der Textvervielfachung differenzieren, die allerdings wechselseitig aufeinander bezogen sind und sich in dieser Interdependenz sowohl visuell als auch strukturell auf die digitale Erscheinungsweise der Texte auswirken. Während die eine Form die (Re-)Präsentation in Richtung des analogen Objekts öffnet, ist die andere durch die digitale Darstellung bedingt und ermöglicht gleichzeitig spezifisch digitale Anschlussoperationen.

Als vielfach instanziiertes Modell der Multiplikation scheint sich die Verdoppelung der Textansicht in Form eines Faksimiles sowie begleitender Transkriptionen zu etablieren, welche die parallele Rezeption visueller und sprachlicher Inhalte der literarischen Objekte ermöglicht. Diese Verdoppelung wie sie bspw. die *Faustedition*, das DTA sowie die *Briefe und Texte aus dem intellektuellen Berlin um 1800* bieten, konstituiert nicht nur konzeptuell, sondern ebenfalls im Hinblick auf das Layout digitaler Infrastrukturen einen gewissen Standard, indem die Verdoppelung häufig als Zweiteilung des Bildschirms in eine linke, objektorientierte und eine rechte, sprachorientierte Seite realisiert ist. Vergleichbar mit den typographischen Dispositiven der Druckkultur interagieren also objektbezogene (Re-)Präsentation und darstellungsräumliche Organisation und prägen durch die spezifische hierbei erzeugte Formation die rezipientenseitige Perspektive auf analoge Literatur. Der Buchseite als strukturierender Einheit wird im Rahmen digitaler Infrastrukturen ein alternatives Organisationsmuster an die Seite gestellt, bei dem gemäß der europäischen Lektürerichtung dem digitalen Faksimile die Vorrangstellung gebührt.[26] Diese Strukturierung begegnet dem langjährigen Bedürfnis einer stärker materialorientierten Textrepräsentation und dies gilt ebenso für die Integration verschiedener Textansichten im Rahmen der Transkription, die eine weitere Vervielfachung bedeutet. Auf allen drei genannten Plattformen lässt sich zwischen einer diplomatischen Umschrift sowie (mindestens) einer „Lesefassung“ umschalten.[27]

Die zweite, genuin digitale Form der Multiplikation, der in noch höherem Maße ein Standardisierungspotential innewohnt, ist die Duplizierung des Textes im Zuge seiner Überführung in ein digitales Datenformat, die zu einer Aufspaltung von Datensammlung und Darstellung für die Rezipientin führt. Die Notwendigkeit zur Standardisierung ergibt sich hier zum einen aus den Maßgaben der Maschinenlesbarkeit und zum anderen aus den vorgesehenen Nutzungsszenarien. Dies verweist auf die veränderten medialen und wissenschaftlichen Bedingungen, an denen sich Archive und Editionen aufgrund ihrer zentralen Funktionen des Sammelns, Bewahrens, Erschließens und Zur Verfügung Stellens orientieren. Ermöglicht eine Überführung in digitale Formate (zumindest potentiell) eine platzsparende und gegenüber Umwelteinflüssen resistentere Bewahrung,[28] so erfordern die Verknüpfung im Sinne von Linked Open Data, die digitale Archivierung sowie die computergestützte Analyse der Texte gleichzeitig eine stärkere Orientierung an allgemeingültigen Standards, welche die (Re-)Präsentation der Texte prägen und sich auf deren Wahrnehmung in der wissenschaftlichen Benutzung auswirken.

[26] Eine solche „Umkehr der funktionalen Relation von Dokument und Text“ fordert für die „elektronische Edition“ bereits Gabler 2006, Abschn. 17, sieht diese zu jenem Zeitpunkt aber als noch nicht realisiert an.

[27] „Lesefassung“ ist der von BTB verwendete Ausdruck. Die *Faustedition* nennt diese „textuelle Transkription“. Das DTA bietet ebenfalls eine „Text-Ansicht“ sowie zwei unterschiedliche Dateiformate (HTML und CAB) mit normierten Zeichen.

[28] vgl. hierzu zurecht einschränkend allerdings Lukas 2013, S. 34.

Hierbei lassen sich allerdings wiederum eine externe und eine interne Ebene der Standardisierung unterscheiden, die sich tendenziell mit einem archivalisch/bibliografischen Ordnungssystem bzw. einer editorisch-philologischen Perspektive verbinden. Erstens erfolgt die Aufnahme von Metadaten verschiedenen archivalischen und bibliografischen Standards; Personen, Werke und Orte werden zudem zunehmend mit zentralen Normdaten verlinkt. Zwar handelt es sich bspw. bei den RNA-Standards,[29] anhand derer die Dokumente des *Handke online*-Portals erfasst sind, zunächst v. a. um eine Überführung analoger Katalogisierungsstandards von Archiven in ein digitales Format und Ähnliches gilt für die Strukturierung anhand der FRBR[30] bei *Ernst Jandl Online* oder die Verlinkung von Registern mit der Gemeinsamen Normdatei (GND) der Deutschen Nationalbibliothek[31] bspw. im DTA und in der *Sturm*-Edition. Allerdings sind zum einen auch diese Übertragungen aufgrund der spezifischen digitalen Anforderungen der Formalisierung als Modellierungsprozesse mit großem Einfluss auf die literaturwissenschaftliche Wissensordnung zu begreifen. Über diese Standards verorten sich die Texte bzw. die in ihnen adressierten Inhalte in zentral konfigurierten Netzwerken, deren Konzeptionierung vonseiten einzelner Projekte kaum beeinflusst werden kann. Zum anderen und damit in Zusammenhang prägen archivalische und bibliografische Standards, die vormals maßgeblich der internen Organisation archivarischer Einrichtungen dienten, in digitalen Infrastrukturen in weit höherem Maße die Wahrnehmung von Literatur im Sinne verbindlicher Zuordnungen, indem sie die Erscheinungsweise von Texten systematisch für die allgemeine Benutzerschaft vorstrukturieren.

Die gesteigerte Notwendigkeit zur Standardisierung im Rahmen der digitalen (Re-)Präsentation besteht nicht nur für die Metadaten, sondern zweitens gleichermaßen für die Texte selbst. Sie begründet trotz anhaltender Kritik an den damit verbundenen Einschränkungen[32] die Vorrangstellung der TEI bei der Repräsentation geisteswissenschaftlicher Texte: Der XML-Dialekt garantiert aktuell das Höchstmaß an Interoperabilität und damit sowohl Nachhaltigkeit als auch Nachnutzbarkeit der digital zur Verfügung gestellten Texte.
Die Standardisierung der philologisch-editorischen Recodierung impliziert hierbei im Gegensatz zu den archivalisch-bibliografischen Standards ein erhöhtes Pluralisierungspotential, indem das TEI-Markup individuelle Erweiterungen sowie verschiedene Perspektivierungen von Text ermöglicht,[33] die wiederum die Grundlage verschiedener Präsentationsformen bilden können. Letzteres gilt nicht nur im Hinblick auf die unterschiedlichen präsentierten Versionen eines Textes, bei denen etwa zwischen diplomatischen Umschriften, verschiedenen Fassungen, einem Lesetext oder der Einblendung von Varianten gewählt werden kann. Zusätzlich ist diese Vervielfältigung in den meisten Fällen an der ‚Textoberfläche' der Infrastruktur für die Rezipientin sichtbar und generiert damit eine weitere Schicht, in der alle editorischen Manipulationen für die Rezipientin nachvollziehbar werden.

Weiterhin unterstützt die ‚Verdatung' Visualisierungen in Form von schematischen grafischen Darstellungen. Diesen Visualisierungen, die entweder auf der Basis der TEI-Annotationen oder

[29] „Regeln zur Erschließung von Nachlässen und Autographen" der Staatsbibliothek zu Berlin sowie der Österreichischen Nationalbibliothek, vgl. http://kalliope-verbund.info/_Resources/Persistent/5bf5cd96ea4448bfec20caf2e3d3063344d76b58/rna-berlin-wien-mastercopy-08-02-2010.pdf.

[30] „Functional Requirements for Bibliographic Records" der International Federation of Library Associations and Institutions, vgl. https://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf.

[31] vgl. http://www.dnb.de/DE/Standardisierung/GND/gnd_node.html.

[32] vgl. u. a. Jannidis 2010, S. 546; Sahle 2013, Kap. 2.4; Pierazzo 2015, S. 119f. Vgl. hierzu aber auch die von der TEI vorgeschlagenen Lösungsansätze unter http://tei-c.org/release/doc/tei-p5-doc/en/html/NH.html.

[33] vgl. Sahle 2013, Kap. 4.2.

direkt auf Grundlage des UTF-8 codierten Textes generiert werden, bergen ein großes epistemisches Potential, indem sie alternative Perspektiven auf die Texte eröffnen und im Zeigen Strukturen explizit sichtbar machen, die anderenfalls nur im Rahmen sehr spezifischer Analysen erfassbar wären.[34] Eben darin besteht gleichzeitig der operative Gehalt, die „kreative, explorative und kognitive Rolle" jener Visualisierungen,[35] was nicht zuletzt durch die Vielfalt der Möglichkeiten zur visuellen Aufbereitung illustriert wird. Erscheinen grafische Darstellungen als paradigmatisch für die Aufbereitung von Inhalten digitaler Infrastrukturen, hängen die konkreten Inhalte, auf die sich jene beziehen sowie die grafischen Ausgestaltungen von den Funktionsweisen der Infrastruktur sowie den Erkenntnisinteressen ab, welche die jeweilige Textsammlung bedienen soll. Dies verdeutlichen beispielhaft die Wortwolken-Ansicht linguistischer Strukturen im DTA, die Visualisierungen von Figurennetzwerken in der digitalen Edition der *historisch-kritischen Ausgabe der Texte Ödön von Horváths* sowie die schematische Visualisierung von Lagenstrukturen in der *Faustedition*. Diese Formen der Textpluralisierung treten vielfach an die Stelle klassischer Kommentarstrukturen, die im Rahmen digitaler Infrastrukturen an Bedeutung verlieren. Sie üben dabei als „Denkzeuge und Erkenntnismittel"[36] eine der Kommentierung vergleichbare interpretative Wirkung auf die (re)präsentierten Texte aus. Jenseits der auf diese Weise gegebenen Pluralisierungsmöglichkeiten in der Text-(Re-)Präsentation haben jene Visualisierungen das Potential, als spezifisch digitale Darstellungsformen dispositiven Charakter auszubilden. Dies wird zum einen durch die Konvergenz der Perspektivierungen mit jenen digitaler Analysen befördert. Zum anderen trägt hierzu die automatisierte Erzeugung von Visualisierungen mittels externer Werkzeuge wie *Voyant*, etwa im DTA und in der *TextGrid*-Bibliothek, bei.
Zudem wirken sowohl Annotationsstandards großer Projekte wie des DTA oder der digitalen Bibliothek des *TextGrid Repository* als auch antizipierte Nutzungsszenarien der digitalen Analyse dem festgestellten Pluralisierungspotential entgegen. Um die Vorteile der digitalen Verfügbarkeit wissenschaftlich und darstellungstechnisch nutzen zu können, müssen möglichst große Korpora nach den gleichen Richtlinien annotiert sein, damit sie computergestützt analysiert und/oder systematisch untereinander verlinkt werden können.[37]

Dies können Korpora von Werken einer einzelnen Autorin sein, wobei die Standardisierung eine Perspektivierung als Gesamtwerk stützt. Ebenso kann sich die Standardisierung aber auch auf ein Korpus von Werken unterschiedlicher Autorinnen beziehen, die dann wiederum im Sinne eines Epochen- oder Gattungsprofils in der Annotation vereinheitlicht werden. Indem hierbei stets literaturwissenschaftliche Kategorien in Datenmodelle übersetzt werden, gehen die Standardisierung auf Ebene der Auszeichnung und jene auf Ebene der literaturwissenschaftlichen Perspektive Hand in Hand.

In Bezug auf beide dargestellten Formen der Textvervielfachung ist die jeweils festgestellte Tendenz zur Standardisierung maßgeblich als Phänomen auf medialer Ebene zu verorten, welches sich zum einen an eine grundlegend veränderte Auffassung von Literatur im Zuge des

[34] vgl. zum Verhältnis von Sagen und Zeigen in der Rezeption von Notationen Krämer/Cancik-Kirschbaum/Trotzke 2012, S. 25.
[35] Krämer/Cancik-Kirschbaum/Trotzke 2012, S. 20; vgl. im Zusammenhang mit der „interpretive nature" durch die sich insbesondere geisteswissenschaftliches Wissen auszeichnet auch Drucker 2014, S. 136.
[36] Krämer 2014, S. 14.
[37] vgl. hierzu bspw. die „Richtlinien zur Texterfassung" des DTA (http://www.deutschestextarchiv.de/doku/basisformat/transkription.html) sowie Projekte wie CorrespSearch (https://correspsearch.net/index.xql?l=de) und Postdata (http://postdata.linhd.es), die auf eine Verlinkung von Materialien aus unterschiedlichen Projekten zielen.

Material Turn knüpft[38] und zum anderen den spezifischen Möglichkeiten digitaler Medialität Rechnung trägt. Die Disposition digitaler Medialität macht die Realisierung eines Konzepts unproblematisch, welches in analogen Medien nur eingeschränkt bzw. unter gesteigerten materiellen und finanziellen Aufwendungen möglich ist,[39] und gleichzeitig eine andere Form der Textaufbereitung notwendig.

Trotz der dargestellten Tendenz zur Standardisierung in der (Re-)Präsentation größerer Korpora lassen sich allerdings hinsichtlich der konkreten Umsetzungen signifikante Variationen feststellen. Diese führen zu Differenzierungen in Bezug auf die Kategorie des literarischen Textes, indem sich literaturwissenschaftliche Perspektiven auf der Ebene der für die Rezipientin visuell erfassbaren Merkmale der Texte manifestieren. Ganz grundlegend zeigt sich dies anhand der jeweiligen Wahl der Basis für die objektorientierte (Re-)Präsentation.

5. Literarische ‚Originale'

Literarische Texte haben kein ‚Original' in dem Sinne, in dem dies für bildkünstlerische Werke veranschlagt werden kann. Diese auf Lessings *Laokoon* rekurrierende Feststellung hat lange Zeit als Basis einer phonozentrischen Literaturwissenschaft gedient. Die digitale Umsetzung der thematisierten Perspektivverschiebung hin zu einer stärkeren Berücksichtigung der materiellen Ebene stärkt zwar den materiellen Status von Literatur, lässt dabei allerdings die Divergenz von Literatur und Bildender Kunst im Hinblick auf deren materielle Existenz deutlicher als zuvor hervortreten. Bestehen nun im Digitalen gesteigerte Möglichkeiten der Abbildung der visuellen Dimension von Literatur, schließt sich daran die Frage an, worin genau diese besteht, welches also die materielle Existenz des literarischen Textes ist, die in ihrer Visualität digital (re)präsentiert werden soll.
Digitale Editionen weisen hierbei grundsätzlich eine Tendenz zur Visualisierung handschriftlich überlieferter Artefakte auf, indem sie oft bereits in ihrer Anlage explizit auf solche fokussiert sind. Dies gilt für solche Texte, die bislang noch nicht gedruckt wurden wie etwa die *Notizbücher Theodor Fontanes* sowie viele der Dokumente der *Briefe und Texte Berliner Intellektueller*. Die (Re-)Präsentation mittels digitaler Faksimilierung begründet sich hier aus den Spezifika des Materials selbst. Insbesondere für Notizbücher und ähnlich inhaltlich disparate Konvolute, so die Annahme, bieten digitale Infrastrukturen Möglichkeiten einer (Re-)Präsentation, die der „Materialität, Medialität und Prozessualität" der Textsorte in höherem Maße als das gedruckte Buch gerecht wird.[40] Auch die *Faustedition* präsentiert allerdings Faksimiles ausschließlich von den handschriftlich überlieferten Materialien zum *Faust*, wohingegen die historischen Druckausgaben lediglich in transkribierter Form auf der Plattform abrufbar sind.[41] Ein Blick in Editionen von Texten neuerer Autoren zeigt allerdings, dass diese Fokussierung auf handschriftliches Material (auch) durch die Entstehungszeit der Konvolute bedingt ist. So werden bspw. auf *Handke online* ebenfalls die maschinenschriftlichen Notate Handkes faksimiliert. Die Möglichkeiten der Visualisierung werden – so lässt sich schließen – also v. a. zur (Re-

[38] vgl. hierzu Lukas 2013, S. 38–40.
[39] vgl. Gabler 2006 sowie in Bezug auf Notizbuch-Editionen auch Radecke 2013, S. 152ff.
[40] Radecke 2013, S. 155; vgl. auch Gabler 2010, S. 50.
[41] Ähnliches gilt auch für die *Briefe und Texte aus dem intellektuellen Berlin*, wo zu den bereits edierten Texten wie bspw. Chamissos *Peter Schlemiels wundersamer Geschichte* ebenfalls nur die handschriftlichen Texte faksimiliert werden, selbst wenn es sich dabei (nur) um Abschriften handelt (https://www.berliner-intellektuelle.eu/manuscript?Schlemiel+de#1).

)Präsentation ‚auktorialer Materialien' genutzt, denen auf diese Weise eine der bildkünstlerischen Perspektive vergleichbare Aura der Originalität verliehen wird.

Dieser textgenetische Ansatz, bei dem der Schwerpunkt auf der Textentstehung liegt, von der insbesondere auch die Notizbücher Zeugnis ablegen,[42] wird auf der Plattform *Musil Online* allerdings dahingehend erweitert, dass nicht nur die Handschriften, sondern ebenfalls die Erstausgaben sowie Teil- und Vorabdrucke in Zeitungen und Zeitschriften als Faksimiles abrufbar sind. Dies ermöglicht es der Rezipientin, nicht nur die Entstehung, sondern ebenfalls die publizierten Erscheinungsformen des *Mannes ohne Eigenschaften* digital nachzuvollziehen.[43] Die historischen Ausgaben erlangen auf diese Weise in der digitalen (Re-)Präsentation ebenfalls Artefakt-Charakter. Musils Werk ist hier nicht nur in seiner inhaltlich-sprachlichen Prozessualität (re)präsentiert, sondern erscheint auch in seinen medialen Formaten als dynamisches Aggregat vielfältiger Aktualisierungsformen, in das sich die digitale Darbietung als (aktuell) letzte Stufe einreiht. Die unterschiedlichen Erscheinungsformen werden auf diese Weise für die Rezipientin unmittelbar anschaulich und können so in ihren spezifischen semantisierenden Qualitäten reflektiert werden.
Im Gegensatz dazu werden im DTA ausschließlich die gedruckten Erstausgaben der Texte faksimiliert. Diese Handhabe steht zwar der editorischen Bevorzugung ‚auktorialer Materialien' entgegen, bewirkt aber gleichzeitig ebenfalls eine Normierung der Perspektive: Die Drucke in der DTA-Ansicht bilden eine neue Generation von „immutable mobiles",[44] indem sie in ihrer historischen Erscheinungsweise stabilisiert und so wiederum dem Status von Originalwerken angenähert werden, die aber gleichzeitig ortsunabhängig aufgerufen werden können. Entsprechend finden die in den Digitalisaten teilweise sichtbaren handschriftlichen Hinzufügungen, selbst wenn sie vom Autor stammen, in den Transkriptionen keine Berücksichtigung.[45]

Eine Pluralisierung lässt sich auch im Hinblick auf die weitergehende (Re-)Präsentation der materiellen Eigenschaften der Objekte feststellen. Zwar sind digitale Darstellungen leistungsfähiger in Bezug auf die Visualisierung einzelner Seiten, dies gilt aber nicht für die „dritte Dimension des Buches",[46] über die erst Aussagen über materielle Zusammenhänge sowie Benutzungsmodalitäten der jeweiligen Objekte ermöglicht werden. So kritisiert Spoerhase, dass „in den gegenwärtigen digitalen Transformationsprozessen das anspruchsvolle räumliche Arrangement buchförmiger Textualität häufig ignoriert" wird, wodurch „das Buchobjekt buchstäblich verflacht".[47] Dies gilt für die digitalen (Re-)Präsentationen des DTA, bei denen Buchrücken keine Berücksichtigung finden. Hingegen begegnen die Editionen des *Faust* und der Notizbücher Fontanes der notorischen Flachheit digitaler Darstellungen in unterschiedlicher Weise: In der Notizbuch-Edition werden die Buchrücken ebenfalls faksimiliert und der Darstellung der Einzelseiten am Ende hinzugefügt. Die *Faustedition* bietet dagegen eine schematische

[42] „Insbesondere weil Notizbücher eine individuelle Arbeitsweise des Dichters belegen, rückten sie ins Zentrum wissenschaftlicher Untersuchungen und Editionskonzepte". Radecke 2013, S. 152.
[43] Allerdings bestehen aktuell keine Verknüpfungen zwischen den einzelnen Segmenten der Text-(Re-)Präsentation. Der Darstellung der Projektziele ist aber zu entnehmen, dass diese in der Fortführung des Projekts angedacht ist.
[44] Latour 1990.
[45] vgl. z. B. die hs. Widmung auf dem Titelblatt von Friedrich Christoph Perthes: *Der deutsche Buchhandel als Bedingung des Daseyns einer deutschen Literatur* (1816; http://www.deutschestextarchiv.de/book/view/perthes_buchhandel_1816?p=7).
[46] Spoerhase 2016, S. 53.
[47] Spoerhase 2016, S. 52f.

Visualisierung der Lagenstrukturen der jeweiligen Objekte, welche die materielle Zusammensetzung der Objekte deutlich differenzierter aufschlüsseln und insbesondere bei schmaleren Konvoluten eine verstärkte Sichtbarkeit der dritten Dimension herstellen.

Anhand dieser Beispiele wird deutlich, dass die Vervielfachung der Textansichten im Digitalen keine vollumfängliche (Re-)Präsentation der gesammelten Texte bewirkt. Eine solche ist aufgrund des dynamischen Charakters schriftlicher Äußerungen, die immer zwischen materieller Existenz und immaterieller ‚Bedeutung', zwischen reaktualisierbarer sprachlicher und fixierter schriftlicher Form changieren, unerreichbar. Vielmehr zeigt sich eine Tendenz zur Auratisierung bestimmter Textzustände, wobei sich das darin zum Ausdruck kommende „Nachlassbewusstsein"[48] nicht allein auf Autorinnen, sondern ebenso auf historisch spezifische mediale Zustände beziehen kann.

Dies trägt insbesondere in der Zusammenschau mehrerer Plattformen zur Perspektiven-Pluralisierung auf der Basis unterschiedlicher (Re-)Präsentationen bei: So erscheint der *Faust* im DTA zwar als stabiles Werk in Form der Cotta'schen Erstausgabe, die *Faustedition* stellt aber dagegen Goethes Arbeitsprozess in den Vordergrund, indem das Werk als Summe vielfältiger Ansätze und Überarbeitungen (re)konstruiert wird. Der gedruckte Text stellt hier nicht mehr die Endstufe eines Produktionsprozesses dar, dessen hand- und maschinenschriftliche Vorstufen zwangsläufig zu Derivaten des endgültigen Produkts degradiert werden müssen. Ebenso wird aber die textgenetische Perspektive durch den historischen Druck im DTA ergänzt und in ihrem Originalitätsstatus relativiert. Indem eine solche Zusammenschau verschiedener (Re-)Präsentationen bei digitalen Plattformen ohne Zugangsbeschränkungen sehr einfach realisierbar ist und durch explizite Verlinkungen noch befördert werden könnte, lässt sich die Perspektivvielfalt nutzerseitig zusätzlich steigern. Der Bildschirm der Nutzerin wird dann zum vereinheitlichenden Rahmen verschiedener Sichtweisen auf einen Text. Allerdings tritt auf diese Weise auch die Frage danach, was ein (literarischer) Text ist und worauf sich die literaturwissenschaftliche Forschung im Kern zu beziehen hat, in ihrer Abhängigkeit von literaturwissenschaftlichen Konzepten sowie produktionspraktischen Bedingungen verstärkt zu Tage.

6. Analoge und digitale Lektüremodi

Durch die im digitalen ‚Intermedium' gesteigerten Möglichkeiten der visuellen Darstellung besteht im Rahmen digitaler Infrastrukturen eine starke Tendenz dahingehend, literarische Texte in ihrer Objekthaftigkeit als konkrete Artefakte zu (re)präsentieren und auf diese Weise neben dem Lesen der Texte detaillierte Betrachtungen von deren visuell erfassbaren Qualitäten zu ermöglichen. Jene Tendenz geht – unabhängig von der Frage, an welche konkreten Textzustände sich die (Re-)Präsentation jeweils knüpft – ebenfalls häufig mit gewissen ‚mimetischen' Anpassungen der digitalen Infrastrukturen an die räumlichen Gegebenheiten und Handhabungen der analogen Ausgangsobjekte einher: Die Einzelseiten bilden weiterhin das begrenzende Moment für die Menge an jeweils gleichzeitig dargestelltem Text; Faksimiles von Vorder- und Rückeinband, vom Schmutztitel sowie potentiell von weiteren weißen Seiten sind den Texten vor- bzw. nachgeschaltet und die Seiten müssen jeweils einzeln ‚umgeblättert' werden. Damit erscheinen die digitalen (Re-)Präsentationen der Texte nicht nur visuell, sondern auch in der

[48] Sina/Spoerhase 2017. Vgl. zur aktuellen Tendenz der „Auratisierung der Archivalien" im Zuge der „Neuentdeckung [...] der Archivarbeit" insbesondere die Einleitung des Bandes, ebd., S. 7–17, hier S. 7 sowie im Zusammenhang mit einem Veränderungsprozess im „Selbstverständnis" der Archive auch Lukas 2013, S. 32.

Benutzung aus dem Paratext der Webseite, in die sie eingebettet sind, herausgehoben, da jene maßgeblich durch die Merkmale digitaler Schrifträumlichkeit im Internet geprägt ist. Diese Paratexte haben v. a. nach unten keine ‚natürliche' Begrenzung, verschiedene Bereiche und Text-Zugriffe werden durch Anklicken von Menüpunkten unmittelbar adressierbar[49] und der Textumfang lässt sich aufgrund der Hypertextstruktur nicht auf den ersten Blick bestimmen. Der durch das Layout exponierte Status, der sich ausschließlich auf den literarischen Text an sich bezieht, unterstreicht dessen Rolle als ‚Original', das es in seiner Erscheinungsweise so genau wie möglich abzubilden gilt.

Die Bewahrung der Strukturen der Ausgangsobjekte steht dabei in einer gewissen Spannung zu den Transformationen, die in adäquateren Darstellungsformen und erweiterten Benutzungsweisen resultieren und somit die Remediatisierung[50] überhaupt erst bedingen bzw. rechtfertigen.[51] In diesem für mediale Übersetzungsprozesse charakteristischen Wechselspiel von Bewahren und Transformieren werden „die Grenzen einzelner Medien ebenso etablier[t] wie unterlaufen".[52]

So kontrastiert bereits die Volltextsuche das kontinuierliche Blättern, indem hier eine ganz andere Lektürepraktik befördert wird. Die Möglichkeit, digitale Faksimiles heranzuzoomen oder – wie in der *Faustedition* – das Transkript direkt auf dem Faksimile einzublenden, verweisen zudem wiederum auf die konstruktivistischen Aspekte der digitalen Darstellungen. Das gilt – wie bereits festgestellt – ebenso für den Einbezug von abstrakt-schematischen Visualisierungen. Diese komplementieren die Objektorientierung der digitalen Infrastrukturen, indem sie als spezifisch digitale Darstellungsformen andere Praktiken der Texthandhabung erfordern. Gleichzeitig befördern sie wiederum einen verstärkt schriftbildlichen Lektüremodus.[53] Analoge und digitale Umgangsformen und Lektüremodi[54] interagieren in der Benutzung digitaler Textsammlungen, was sich auch in den ‚kinetischen Qualitäten'[55] der Plattformen niederschlägt: Zum einen werden analoge und digitale Modi auf dieser Ebene einander angenähert, indem sich die Handhabung nutzerseitig ausschließlich auf das Klicken, Scrollen und Zoomen beschränkt. Zum anderen verhindert das analoge Moment des Blätterns in Bezug auf die Handhabung der literarischen Texte die digitale Praxis des Scrollens.

Jenseits dieses Spannungsfeldes übersetzen einige digitale Infrastrukturen die literarischen Texte aber auch vollständig in das mediale Dispositiv der Webseite. Dies gilt bspw. für die digitale *Horváth-Edition* sowie die digitale Bibliothek von *TextGrid*. Beide Plattformen bieten keine Faksimiles, sondern die Texte werden als digitaler Endlostext ohne fixierte untere Seitenbegrenzung dargeboten. Insbesondere im Fall der *TextGrid*-Bibliothek geht hiermit eben-

[49] Krämer 2012, S. 22f. verweist mit Bezug auf Sandbothe darauf, dass sich beim „Schriftphänomen ‚Link'" „[a]nders als bei den vertrauten Fußnoten [...] nicht der Nutzer selbst bewegen [muss] – und sei es nur durch den ‚visuellen Gang' ans Ende der Seite [...] – sondern diese Bewegung macht für ihn der Link." Allerdings bewegt sich die Nutzerin auch in Bezug auf den Link durchaus. Nur gilt dies nicht für die Augen, sondern für die Hand, welche dafür genutzt werden muss, den Link zu aktivieren.

[50] vgl. zu diesem Terminus Bolter/Grusin 2000.

[51] vgl. z. B. Gabler 2010, S. 46.

[52] Benthien/Klein 2017, S. 11.

[53] vgl. wiederum Krämer/Cancik-Kirschbaum/Trotzke 2012, S. 25.

[54] vgl. hierzu allgemein auch Vandendorpe 2013.

[55] Der Terminus der ‚kinetischen' Qualitäten erscheint im Hinblick auf digitale Infrastrukturen angemessener als jener der haptischen Qualitäten, von denen gemeinhin in Bezug auf Manuskripte und Drucke gesprochen wird, weil die Möglichkeiten, die die Benutzerin hat, um die digital (re)präsentierten Objekte im Rahmen der Infrastruktur zu bewegen, sich nicht zwangsläufig an differente haptische Handhabungsformen knüpfen.

falls eine Angleichung einer großen Sammlung unterschiedlicher Texte im Hinblick auf die visuelle Erscheinungsweise einher: Alle Texte erscheinen in der gleichen Schriftart, -größe und Seitenformatierung. Die bereits für das DTA festgestellte Tendenz zur Standardisierung ist deshalb hier in noch höherem Maße gegeben. Diese (Re-)Präsentationsform spiegelt zudem erneut die auf den germanistischen literaturwissenschaftlichen Kanon ausgerichtete Anlage der *TextGrid*-Bibliothek. Die Fokussierung auf deutschsprachige Texte hier wie in der Sammlung des DTA kann auch als Folge antizipierter Nutzungsszenarien verstanden werden, da Mehrsprachigkeit im Rahmen computergestützter Analysen nach wie vor eine große Herausforderung darstellt.

Die als digitale Erweiterung der analogen historisch-kritischen Ausgabe konzipierte *Horváth-Edition* bietet hingegen aufgrund der autorzentrierten Perspektive deutlich mehr Möglichkeiten zur Individualisierung der Text-(Re-)Präsentation. Im Zuge der medialen Transformation, bei der materiell-räumliche Merkmale der Ausgangsobjekte unberücksichtigt bleiben,[56] gewinnen literarische Strukturen in der texträumlichen Organisation an Relevanz. So sind die beiden abrufbaren Ansichten („Leseansicht“ und „Netzwerkansicht“) nach „Bildern“ und „Szenen“, den von Horváth für seine Bühnenstücke etablierten Segmentierungen, gegliedert.[57] Wiederum lässt sich also feststellen, dass sich im Zuge der digitalen (Re-)Präsentation analoger literarischer Schriftlichkeit zwei grundsätzliche Tendenzen gegenüberstehen, die gleichzeitig spezifische literaturwissenschaftliche Schwerpunktsetzungen hinsichtlich der Frage, was literarische Textualität ausmacht, implizieren: Die ‚mimetische‘ Remediatisierung betont den Artefakt-Charakter der Texte. Sie fixiert dabei spezifische mediale Erscheinungsformen als Referenzpunkte der digitalen Rezeption und stellt diese gleichrangig neben die sprachliche Gestalt der Texte.[58] Demgegenüber setzt sich in der Überführung der Texte in ein digitales Dispositiv die für die analoge (Re-)Präsentation etablierte Recodierungskette fort, indem handschriftlich oder gedruckt vorliegende Texte an die Merkmale digitaler Schriftlichkeit angepasst werden. Der Schwerpunkt liegt hierbei also auf den sprachlichen Eigenschaften, wohingegen die digitale Erscheinungsform im Sinne des „Paradigma[s] des immateriellen Textes“[59] vorangegangene mediale Aggregatszustände vollständig ersetzt.

Dass diese beiden Tendenzen nicht im Sinne einer starren Opposition, sondern vielmehr als Pole einer Skala möglicher (Re-)Präsentationsformen aufzufassen sind, verdeutlichen Mischformen analoger und digitaler Schriftlichkeitsmerkmale im Rahmen einer Infrastruktur, die wiederum Anhaltspunkte für variierende Konzeptionierungen von Textualität bieten: So reproduziert die „dokumentarische Transkription“ der *Faustedition* ebenfalls das analoge Merkmal des Blätterns, wohingegen die parallel einblendbare „textuelle Transkription“ gescrollt wird und aufgrund ihrer Erweiterbarkeit durch die jeweiligen Varianten zu jeder Textzeile von variablem Umfang ist.

---

[56] Interessant ist in diesem Zusammenhang das gegensätzlich konzipierte Zusammenspiel von Print-Edition und digitaler Plattform in der Horváth-Edition und der *Faustedition*: Während letztere als Ergänzung zur digitalen Version analog einen konstituierten Lesetext bereitstellen wird (vgl. http://beta.faustedition.net/intro), ist ein eben solcher die Basis der digitalen Horváth-Edition, wohingegen in der analogen Edition Textgenese, Faksimiles sowie Rekonstruktionen der Montage-Technik Horváths geboten werden (vgl. http://gams.uni-graz.at/context:ohad/sdef:Context/get?locale=ueber&context=analog).

[57] Im Gegensatz dazu orientiert sich die Gliederung der zugehörigen Print-Edition vorrangig an textgenetischen Merkmalen (vgl. hierzu bspw. das Vorwort in Horváth 2009, S. 1–14).

[58] In erhöhtem Maße gilt dies für solche Plattformen wie etwa *Handke online*, die überhaupt keine Transkriptionen enthalten, sondern die digitalen Faksimiles in ausschließlich kontextualisierende Paratexte einbetten.

[59] Lukas/Nutt-Kofoth/Podewski 2014, S. 1.

Die Interferenz digitaler und analoger Lektüremodi verweist daher auch auf die komplexe Verschränkung texttheoretischer Perspektiven, antizipierter Nutzungsszenarien und medialer Bedingungen bei der Ausgestaltung der strukturierenden Konzepte digitaler Infrastrukturen.

7. Resümee

Die anhand der digitalen Infrastrukturen zur (Re-)Präsentation analoger literarischer Texte beobachtete Tendenz zur Standardisierung geht maßgeblich aus den spezifischen technischen Bedingungen digitaler Medialität hervor und ist zudem in Interdependenz mit der Überführung archivalischer und bibliografischer Standards in öffentlich zugängliche digitale Räume sowie den aktuellen Entwicklungen im Bereich der digitalen Analyse zu sehen. Indem Suchanfragen in Sammlungen, Darstellungen digital (re)präsentierter Texte und mithilfe digitaler Werkzeuge erzeugte Visualisierungen auf der gleichen Datenbasis operieren, rücken im Rahmen digitaler (Re-)Präsentation nicht nur Archiv und Edition konzeptuell näher zusammen, sondern es bestehen ebenfalls Konvergenzen mit textanalytischen Formaten.
Die wechselseitig aufeinander bezogenen strukturellen und visuellen Textvervielfachungen, welche Archiv, Edition und Analyse verknüpfen, dienen dabei einerseits der Erhöhung von Evidenz und Nachvollziehbarkeit literaturwissenschaftlicher Forschung, indem die „Hin- und Rückverbindung“[60] zwischen abwesendem Objekt und Diskurs über das Objekt gestärkt wird.[61] Digitale Paradigmen der Textaufbereitung in Form von TEI-Codierungen, Figurennetzwerken oder Wortwolken befördern andererseits in unterschiedlichen Sichtbarkeitsgraden eine spezifisch an die digitale Medialität geknüpfte Perspektive auf literarische Texte.[62]

Dabei verweisen allerdings die beschriebenen Varianzen hinsichtlich der Korpusbildung, Objektauswahl und Textaufbereitung darauf, wie im Rahmen dieser digitalen (Re-)Präsentationen „epistemische Dinge“[63] in spezifischer Form hergestellt werden. Die festgestellten Ausdifferenzierungen sind hierbei wiederum maßgeblich auf die Möglichkeiten der Vervielfältigung von Textansichten und Benutzungsformen zurückzuführen, die das digitale ‚Intermedium‘ bietet. Die Überwindung der medialen Beschränkungen des Buchformats befördert eine Pluralisierung, indem unterschiedliche analoge Objekte den Ausgangspunkt digitaler ‚Ableitungen‘ bilden[64] und dem – auch digital weiterhin dominanten – Autor-Werk-Paradigma analoger Editionen zunehmend alternative Programme der Zusammenhangstiftung und Kontextualisierung zur Seite gestellt werden.

---

[60] vgl. hierzu die Ausführungen zur Überzeugungskraft gezeichneter Landkarten in Latour 1990, S. 28 (vgl. zur deutschen Übersetzung Latour 2006, S. 268).

[61] „Als Erinnerung verweist ein Faksimile auf das absente Original, verweist damit aber auch aus der Edition hinaus. Als integrierte Vergegenwärtigung hingegen verleiht es dem Original die Kraft der Präsenz.“ Gabler 2006, Abschn. 12.

[62] Zudem gilt hier ebenfalls, was Latour 1990, S. 28 bereits für das Beispiel analoger zentralperspektivischer Darstellungen feststellt: „Impossible palaces can be drawn realistically, but it is also possible to draw possible objects as if they were utopian ones.“

[63] vgl. Rheinberger 2006; vgl. in diesem Zusammenhang auch Martus’ Übertragung auf den Bereich der Philologie (Martus 2015).

[64] Gabler 2006, Abschn. 16 vertritt die Vorstellung, „dass Überliefern und Edieren – und vor allem: Edieren – nichts anderes ist und nie etwas anderes gewesen ist, als die *Konstruktion von Texten* abgeleitet von der Evidenz – und das meint nun dezidiert nicht mehr: der Zeugenschaft, sondern der Anschauung – der Dokumente“ (Hvh. i. O.). Er verkennt dabei allerdings, dass die Auswahl der Dokumente, aus denen abgeleitet wird, ebenfalls einen konstruktiven Prozess impliziert.

Dies verweist auf die verstärkte Relevanz von Selektions- und Strukturierungsparametern im Rahmen der digitalen Aufbereitung in Abhängigkeit von philologischen Erkenntnisinteressen und nutzungsorientierten Erwägungen. Eine vollumfängliche (Re-)Präsentation der jeweiligen Gegenstände ist ebenso wie eine unbegrenzte Anzahl von Perspektivierungen im Digitalen lediglich potentiell erreichbar, erscheint praktisch aber nicht realisierbar. Deshalb begünstigen digitale Formate zwar die Etablierung alternativer Ordnungsformen, diese bedingen allerdings gleichzeitig verstärkte Investitionen in die Entwicklung plausibler Konzepte zu deren Eingrenzung und Strukturierung, die den Anforderungen von Medium, Gegenstand und Nutzerschaft gerecht werden. Die verstärkte Annäherung an das Objekt geht deshalb gleichermaßen mit einer erhöhten konstruktiven Leistung seitens der bearbeitenden Wissenschaftlerin einher.

Die Remediatisierung analoger Literatur in digitaler Form dient der ‚Entautomatisierung' von (Re-)Präsentationsstrukturen literarischer Texte und damit – produktions- wie rezeptionsseitig – der Reflexion der an Literatur als bedeutungstragend wahrgenommenen Charakteristika. Die potentielle Vielfalt digitaler (Re-)Präsentationsmöglichkeiten führt deshalb dazu, dass sowohl die Textauswahl als auch die konkreten Formen der digitalen Aufbereitung verstärkt theoretisch geleitete Konzepte transportieren.

In diesem Zusammenhang bestehen mit der Orientierung am analogen Ausgangsobjekt sowie dem verstärkten Einsatz von grafischen Visualisierungen zur Darstellung textueller Zusammenhänge zwei komplementäre Tendenzen, die allerdings gleichermaßen Textumgangsformen befördern, die sich verstärkt an den schriftbildlichen Charakteristika der (Re-)Präsentationen orientieren. Der auf das Zeigen als epistemische Praktik gerichtete Rezeptionsmodus ist auch deshalb für die untersuchten digitalen Infrastrukturen als besonders maßgeblich anzusehen, weil sich hier Visualisierungsstrategien, die sich an die spezifischen Bedingungen digitaler Schriftlichkeit knüpfen, ein gesteigertes Interesse an der Materialität von Literatur und modernes Nachlassbewusstsein treffen. In ihrer Komplementarität haben sie an den festgestellten Interferenzen zwischen analogen und digitalen Lektüremodi teil, die aktuell die Ausgestaltung digitaler Infrastrukturen prägen.
Die vorangegangenen Untersuchungen haben gezeigt, dass der (Re-)Präsentation analoger literarischer Texte in digitalen Infrastrukturen gleichermaßen ein gegenüber analogen Formaten gesteigertes Potential zur Pluralisierung wie zur Standardisierung innewohnt. Die pluralisierenden Tendenzen stehen dabei in Spannung zur Strukturierung digital (re)präsentierter Texte durch übergeordnete Metadaten-Standards und Annotationsrichtlinien, welche zunehmend literaturwissenschaftliche Episteme hinsichtlich der Wahrnehmung von Literatur mitbestimmen. Die Tatsache, dass die herausgearbeiteten Merkmale der (Re-)Präsentation literarischer Texte im Rahmen digitaler Infrastrukturen weitgehend unabhängig von den konkret aufbereiteten Materialien bestehen, verweist auf einen generellen Prozess der Aushandlung von (Re-)Präsentationsformaten im Spannungsfeld medialer Erfordernisse, literaturwissenschaftlicher Konzepte und Nutzerorientierung, der perspektivisch zu noch stärkeren Angleichungen im Sinne der Ausprägung digitaler Dispositive der (Re-)Präsentation literarischer Texte führen könnte. Diese Suchbewegungen mögen auch ein Grund dafür sein, dass der Disposition zur Vervielfältigung von Textauswahl- und (Re-)Präsentationskriterien eine Tendenz zur Bevorzugung bereits kanonisierter Texte und Autorinnen gegenübersteht. Inwieweit diese Form der Standardisierung lediglich das Resultat einer Übergangsphase darstellt, bleibt abzuwarten.

Literaturverzeichnis

Primärquellen:

Briefe und Texte aus dem intellektuellen Berlin um 1800. Projektverantwortliche: Dr. Anne Baillot, https://www.berliner-intellektuelle.eu (Zugriffsdatum: 15.12.2018).

The complete Writings and Pictures of Dante Gabriel Rossetti, a Hypermedia Archive. Hg. von Gerome J. McGann 2007, http://www.rossettiarchive.org/index.html (Zugriffsdatum: 15.12.2018).

Deutsches Textarchiv. Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache. Hg. von der Berlin-Brandenburgischen Akademie der Wissenschaften 2018, http://www.deutschestextarchiv.de (Zugriffsdatum: 15.12.2018).

Historisch-kritische Faustedition. Hg. von Anne Bohnenkamp / Silke Henke / Fotis Jannidis. Unter Mitarbeit von Gerrit Brüning / Katrin Henzel / Christoph Leijser / Gregor Middell / Dietmar Pravida / Thorsten Vitt / Moritz Wissenbach. Version Beta 3 2017, URL: http://beta.faustedition.net (Zugriffsdatum: 15.12.2018).

Theodor Fontane: Notizbücher. Digitale genetisch-kritische und kommentierte Edition. Hg. von Gabriele Radecke 2018, https://fontane-nb.dariah.eu/index.html (Zugriffsdatum: 15.12.2018).

Handke online. Hg. von der Österreichischen Nationalbibliothek 2015, https://handkeonline.onb.ac.at (Zugriffsdatum: 15.12.2018).

Heinrich-Heine-Portal. Hg. vom Heinrich-Heine-Institut Düsseldorf und Kompetenzzentrum Trier 2009, http://germazope.uni-trier.de:8080/Projekte/HHP/ (Zugriffsdatum: 15.12.2018).

Ödön von Horváth: Wiener Ausgabe sämtlicher Werke, Bd. 4: Kasimir und Karoline. Hg. von Klaus Kastenberger / Kerstin Reimann, Wien 2009.

Ödön von Horváth. Historisch-kritische Ausgabe – Digitale Edition. Hg. von Klaus Kastenberger 2018, http://gams.uni-graz.at/context:ohad (Zugriffsdatum: 15.12.2018).

Ernst Jandl Online. Hg. von Vanessa Hannesschläger / Ludwig Boltzmann Institut f. Geschichte u. Theorie der Biographie und d. Österreichischen Nationalbibliothek 2015, http://jandl.onb.ac.at (Zugriffsdatum: 15.12.2018).

Musil Online. Hg. von Robert-Musil-Institut im Kärntner Literaturarchiv 2016, http://musilonline.at (Zugriffsdatum: 15.12.2018).

Parzival-Projekt. Projektverantwortlicher: Michael Stolz, http://www.parzival.unibe.ch/home.html (Zugriffsdatum: 15.12.2018).

Der Sturm: Digitale Quellenedition zur Geschichte der internationalen Avantgarde. Hg. von Marjam Trautmann / Torsten Schrade, Mainz, Akademie der Wissenschaften und der Literatur, 2018, https://sturm-edition.de (Zugriffsdatum: 15.12.2018).

TextGrid Digitale Bibliothek. Hg. von der Niedersächsischen Staats- und Universitätsbibliothek Göttingen 2018, https://textgrid.de/digitale-bibliothek (Zugriffsdatum: 15.12.2018).

The Walt Whitman Archive. Hg. von Ed Folsom / Kenneth M. Price 2018, https://whitmanarchive.org (Zugriffsdatum: 15.12.2018).

Forschungsliteratur:

Bender, Michael: Forschungsumgebungen in den Digital Humanities. Nutzerbedarf, Wissenstransfer, Textualität. Berlin / Boston: De Gruyter 2016.

Benthien, Claudia / Klein, Claudia: Praktiken des Übersetzens und Rahmens. Zur Einführung. In: Dies (Hg.): Übersetzen und Rahmen: Praktiken medialer Transformation. Paderborn: Fink 2017, S. 9–16.

Bolter, Jay David / Grusin, Richard: Remediation. Understanding New Media. Cambridge / London: MIT Press 2000.

Driscoll, Matthew James / Pierazzo, Elena (Hg.): Digital Scholarly Editing. Theories, Models and Methods. Cambridge: Open Book Publishers 2016.

Drucker, Johanna: Graphesis: Visual Forms of Knowledge Production. Cambridge/MA: Harvard University Press 2014.

Gabler, Hans Walter: Das wissenschaftliche Edieren als Funktion der Dokumente. In: Jahrbuch für Computerphilologie 8. Hg. von Georg Braungart, Peter Gendolla und Fotis Jannidis. Paderborn: mentis 2006. http://computerphilologie.digital-humanities.de/jg06/gabler.html (Zugriffsdatum: 15.12.2018).

Gabler, Hans Walter: Theorizing the Digital Scholarly Edition. In: Literature Compass 7, 2 (2010), S. 43–56.

Gengnagel, Tessa: The ‚Beta Dilemma' – A Review of the Faust Edition. In: ride 7 (2017). http://ride.i-d-e.de/issues/issue-7/faustedition/ (Zugriffsdatum: 15.12.2018).

Jäger, Ludwig: Intermedialität – Intramedialität – Transkriptivität. In: Arnulf Deppermann / Angelika Linke (Hg.): Sprache intermedial. Stimme und Schrift, Bild und Ton. Berlin: De Gruyter 2010, S. 301–324.

Jannidis, Fotis: Digital Editions in the Net. Perspectives for Scholarly Editing in a Digital World. In: Jörgen Schäfer / Peter Gendolla (Hg.): Beyond the Screen. Transformations of Literary Structures, Interfaces and Genres. Bielefeld: Transcript 2010, S. 543–560.

Jockers, Matthew J: Macroanalysis. Digital Methods & Literary History. Urbana u. a.: University of Illinois Press 2013.

Krämer, Sybille / Cancik-Kirschbaum, Eva / Trotzke, Rainer: Einleitung. In: Dies. (Hg.): Schriftbildlichkeit. Wahrnehmbarkeit, Materialität und Operativität von Notationen. Berlin: De Gruyter 2012, S. 13–35.

Krämer, Sybille: Zur Grammatik der Diagrammatik. Eine Annäherung an die Grundlagen des Diagrammgebrauches. In: LiLi. Zeitschrift für Literaturwissenschaft und Linguistik 44, 176 (2014), S. 11–30.

Latour, Bruno: Drawing Things Together. In: Michael E. Lynch / Steve Woolgar (Hg.): Representation in Scientific Practice. Cambridge: MIT Press 1990, S. 19–68.

Latour, Bruno: Drawing Things Together. Die Macht der unveränderlich mobilen Elemente. In: Andréa Belliger / David J. Krieger (Hg.): ANThology: Ein einführendes Handbuch zur Akteur-Netzwerk-Theorie. Bielefeld: Trancsript 2006, S. 259–307.

Lukas, Wolfgang: Was ist das Digitalisierungsinteresse der geistes- und kulturwissenschaftlichen Forschung? In: LVR-Archivberatungs- und Fortbildungszentrum (Hg.): Digital und analog.

Die beiden Archivwelten. 46. Rheinischer Archivtag, Ratingen 21.–22. Juni 2012. Beiträge. Bonn: Habelt 2013, S. 32–47.

Lukas, Wolfgang / Nutt-Kofoth, Rüdiger / Podewski, Madleen: Zur Bedeutung von Materialität und Medialität für Edition und Interpretation. Eine Einführung. In: Dies. (Hg.): Text – Material – Medium. Zur Relevanz editorischer Dokumentationen für die literaturwissenschaftliche Interpretation. Berlin / Boston: De Gruyter 2014, S. 1–22.

Manovich, Lev: The Language of New Media. London: MIT Press 2001.

Martus, Steffen: Epistemische Dinge in den Literaturwissenschaften? In: Andrea Albrecht / Lutz Danneberg / Carlos Spoerhase (Hg.): Theorien, Methoden und Praktiken des Interpretierens. Berlin: De Gruyter 2015, S. 23–51.

Nantke, Julia: New Practices = New Conditions? Interrelations of Practical Approaches, Methodologies and Theoretical Concepts in Digital Literary Studies. In: Pal Kelemen / Nicolas Pethes (Hg.): Philology in the Making. Analog/Digital Practices of Scholarly Reading and Writing. Bielefeld: Transcript 2019, S. 197-217.

Nutt-Kofoth, Rüdiger: Sichten – Perspektiven auf Text. In: Anne Bohnenkamp (Hg.): Medienwandel/Medienwechsel in der Editionswissenschaft. Berlin / Boston: De Gruyter 2013, S. 19–29.

Nutt-Kofoth, Rüdiger: Zum Verhältnis von Nachlasspolitik und Editionskonzeption. In: Kai Sina / Carlos Spoerhase (Hg.): Nachlassbewusstsein. Literatur, Archiv, Philologie 1750–2000. Göttingen: Wallstein 2017, S. 92–111.

Plachta, Bodo: Editionswissenschaft. Stuttgart: Reclam 2006.

Radecke, Gabriele: Notizbuch-Editionen. Zum philologischen Konzept der Genetisch-kritischen und kommentierten Hybrid-Edition von Theodor Fontanes Notizbüchern. In: edito 27 (2013), S. 149–172.

Rheinberger, Hans-Jörg: Experimentalsysteme und epistemische Dinge. Eine Geschichte der Proteinsynthese in Reagenzglas. Frankfurt a. M.: Suhrkamp 2006.

Robinson, Peter: What a text really is not, and why editors have to learn to swim. In: Literary and Linguistic Computing 24, 1 (2009), S. 41–52.

Sahle, Patrick: Digitales Archiv – Digitale Edition. Anmerkungen zur Begriffsklärung. In: Michael Stolz et al (Hg.): Literatur und Literaturwissenschaft auf dem Weg zu den neuen Medien. Eine Standortbestimmung. Zürich: germanictsik.ch 2007, S. 64–84.

Sahle, Patrick: Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Teil 3: Textbegriffe und Recodierung. Norderstedt: BoD 2013.

Sahle, Patrick: What is a Scholarly Digital Edition? In: Matthew James Driscoll / Elena Pierazzo (Hg.): Digital Scholarly Editing: Theories and Practices. Cambridge: Open Book Publishers 2016, S. 19–39.

Sina, Kai / Spoerhase, Carlos (Hg.): Nachlassbewusstsein. Literatur, Archiv, Philologie 1750–2000. Göttingen: Wallstein 2017, S. 92–111.

Spoerhase, Carlos: Linie, Fläche, Raum. Die drei Dimensionen des Buches in der Diskussion der Gegenwart und der Moderne. Göttingen: Wallstein 2016.

Vandendorpe, Christian: Reading on Screen: The New Media Sphere. In: Ray Siemens / Susan Schreibman (Hg.): A Companion to Digital Literary Studies. Oxford u. a.: Wiley-Blackwell 2013, S. 203–215.

Wehde, Susanne: Typographische Kultur. Eine zeichentheoretische und kulturgeschichtliche Studie zur Typographie und ihrer Entwicklung. Tübingen: De Gruyter 2000.

Wizisla, Erdmut: Archive als Editonen? Zum Beispiel Bertolt Brecht. In: Rüdiger Nutt-Kofoth / Podo Plachta / H. T. M. van Vliet / Hermann Zwerschina (Hg.): Text und Edition. Positionen und Perspektiven. Berlin: Erich Schmidt 2000, S. 407–417.

Aristoteles auf Reisen: Handschriftenforschung in der digitalen Infrastruktur des SFB 980 „Episteme in Bewegung“

Michael Krewet, Philipp Hegel, Germaine Götzelmann, Danah Tonne, Sibylle Söring (Berlin)

1. Einleitung

In dem folgenden Beitrag möchten wir an einem Fallbeispiel aus dem Kerngebiet der Klassischen Philologie zeigen, wie eine digitale Infrastruktur neue Möglichkeiten für Forschungsfragen schaffen kann, die bislang als nicht oder nicht innerhalb eines Forscherlebens beantwortbar gelten – auch durch sich eröffnende neue Praktiken der Zusammenarbeit.[1] Von dem Fallbeispiel ausgehend wird am Ende noch anhand des Sonderforschungsbereichs *Episteme in Bewegung* (SFB 980) ein Ausblick gegeben, in welcher Weise eine solche bestehende Infrastruktur für weitere Projekte Perspektiven eröffnen kann.

2. Das Fallbeispiel

Die Werke der griechischen und auch lateinischen Antike wurden über Jahrhunderte hin handschriftlich in Form von Kopien überliefert, zunächst über Papyrusrollen, später über Pergament- oder Papiercodices. Vor allem die Codices sind uns in vielen Fällen erhalten. Diese haben wir in ihrer großen Mehrheit zeitlich in etwa zwischen das 9. und 16. Jahrhundert zu datieren.[2]

Die Erforschung der handschriftlichen Überlieferung einer Schrift (v.a. über Codices) bildet seit jeher ein Kerngebiet der Klassischen Philologie. Einen wichtigen Grund für die Zuwendung zu der Überlieferungsgeschichte stellt die Notwendigkeit der Erforschung von Verwandtschaftsverhältnissen unter den einzelnen Kopien dar, um, wenn diese Verhältnisse geklärt sind, voneinander unabhängigen Lesarten, bzw. Kopien, zu identifizieren. Die Berücksichtigung unabhängiger Lesarten ist wiederum für gute Texteditionen unerlässlich.[3] Die Methode, die hierfür verwendet wird, ist die Textkollation. Der handschriftliche Text einer Kopie wird in der Regel transkribiert. Die transkribierten Texte werden miteinander verglichen und auf unterschiedliche Varianten und Fehler geprüft.

Einen zweiten wichtigen Grund für die Erforschung der Überlieferung bildet der Nachvollzug der räumlichen und zeitlichen Verbreitung einer Schrift. Über paläographische oder auch kodikologische Eigenheiten oder Vermerke können Handschriften bestimmten Zeiten oder auch geographischen Regionen und Orten, teilweise auch einzelnen Kopisten zugewiesen werden.[4] So können z.B. auch Zentren, in denen bestimmte Werke kopiert und vielleicht auch besonders studiert wurden, nachgewiesen werden.

---

[1] Im vorliegenden Beitrag wurde der Stil des Vortrags weitgehend beibehalten.

[2] vgl. hierzu und auch dem Folgenden in der Einleitung umfassend und exemplarisch: Crisci, Edoardo / Degni, Paola: La Scrittura Greca dall’ Antichità All’ Epoca della Stampa. Rom: Carocci 2011.

[3] Ausführlich hierzu West, Martin L.: Textual Criticism and Editorial Technique: applicable to Greek and Latin texts. Stuttgart: Teubner 1973. Ferner: Maas, Paul: Textkritik. Leipzig, Berlin: Teubner 1927.

[4] Ein guter Einblick und Überblick findet sich in den online gestellten Handschriftenbeschreibungen auf der Homepage des CAGB-Projektes der BBAW: https://cagb-db.bbaw.de/register/werke.xql?cRef=Int. (Zugriffsdatum: 10.10.2018).

Einen dritten wichtigen Grund für die Erforschung der Handschriften in Verbindung mit der Überlieferung bildet heute das Interesse an Verwendungskontexten einzelner Handschriften und an wissensgeschichtlichen Eigenheiten der Handschriften. Wurde der Texte für das Studium innerhalb eines gelehrten Kreises o.ä. kopiert? Oder wurde er als ein Bibliotheksexemplar kopiert? Oder für ein privates Studium? Mit diesen Fragen verbindet sich beispielsweise das Interesse an der ‚Layoutgestaltung' einer einzelnen Handschrift: Wenn viel Raum an den Rändern gelassen wurden, damit ausreichend Platz für Glossen, Scholien, Diagramme oder Kommentare vorhanden war, so liegt die Wahrscheinlichkeit höher, dass es sich um ein Exemplar für das Studium handelt; wenn der Rand geringer war, ist es wahrscheinlicher, dass es sich um ein Bibliotheksexemplar handelt. Wenn die Schrift informell war, liegt es wiederum näher, dass es sich um ein Exemplar für den privaten Gebrauch (oder auch das Studium in einer gelehrten Gruppe) handelt. Wenn es sich um eine formale Schönschrift handelt, liegt es näher, dass es sich um ein Bibliotheksexemplar handelt usw. Damit verbinden sich Fragen, wo und wann es gelehrte Kreise gab, die sich dem Werk widmeten, oder in welchen Bibliotheken oder auch Regionen zu welcher Zeit ein Exemplar einer Schrift verfügbar war. Eine weitere Fragestellung ist angesichts zahlreicher anonymer Kommentare an den großen Rändern der Handschriften etwa, ob sich das Philosophieren über Jahrhunderte hin in bestimmten Kreisen als Kommentieren verstand. Gab es ferner didaktische Strategien in der Aufbereitung eines Textes einer Handschrift für ein besseres Verständnis? Welches Wissen wurde in welcher Form zur Kommentierung an den Text in Form von Glossen, Scholien, Kommentaren oder Diagrammen herangetragen und war damit zu bestimmten Zeiten an bestimmten Orten verfügbar? Und wie veränderte die Art der Kommentierung oder auch didaktischen Aufbereitung des Textes das Verständnis der Schrift?

3. Das Beispiel einer konkreten Forschungsfrage

Im Fall unseres Beitrags möchten wir aus diesem Komplex nur eine einzelne Forschungsfrage am Beispiel einer einzelnen antiken Schrift – nämlich Aristoteles' Schrift *de interpretatione* – herausgreifen, die perspektivisch die großen Vorteile einer digitalen Forschungsinfrastruktur für die Bearbeitung lange bestehender Forschungsfragen aufzeigen kann. Von Aristoteles' Schrift *de interpretatione* sind uns ca. 150 griechischsprachige Handschriften erhalten.[5] Dies ist vergleichsweise eine extrem hohe Zahl für ein griechischsprachiges Werk der Antike.

Das Beispiel dieser Schrift des Aristoteles bildet einen Fall, in dem die Erforschbarkeit der Überlieferungsgeschichte auf ‚analogem' Weg an ihre Grenzen stößt.[6] Führende Textwissenschaftler der vergangenen Jahrzehnte haben die Schwierigkeit bzw. die Unerforschbarkeit der Überlieferungsgeschichte dieser Schrift betont. Den Grund sehen sie in der Komplexität, in der sich die Überlieferung darbietet.[7] Die Grundannahme der Textkritik, die immer wieder eine

[5] vgl. zu allgemeinen Charakteristika der handschriftlichen Überlieferung von *de interpretatione*: Montanari, Elio: La sezione linguistica del *Peri Hermeneias* di Aristotele, 2 Bde. (Studi e testi 5) Florenz: Università degli studi die Firenze, Dipartimento di scienze dell'antichità "Giorgio Pasquali" 1984.

[6] Ausführlich ebd., Bd. 1, S. 49ff.

[7] vgl. neben Montanari 1984 v.a.: Minio Paluello, Lorenzo (Hg.): Aristotelis Categoriae et Liber de interpretatione. Oxford: Clarendon 1949, hier v.a.: Praefatio XIX; Reinsch, Diether Roderich: Fragmente einer Organon-Handschrift des zehnten Jahrhunderts aus dem Katharinenkloster auf dem Berg Sinai. In: Philologus 145,1 (2001), S. 57-69, und Weidemann, Hermann (Hg.): Aristoteles, de interpretatione. Berlin, New York: De Gruyter 2014, Praefatio XXVI.

empirische Bestätigung findet, lautet (etwas vereinfacht), dass beim Kopieren eines Werkes Fehler passieren (bzw. allgemeiner und neutraler: Lesarten entstehen). Eine Kopie enthält dieser Grundannahme zufolge immer mehr Fehler im Vergleich zu ihrer Vorlage. Oder anders formuliert: Sie enthält die Fehler, die auch ihre Vorlage enthält, und ferner zusätzliche Fehler. Über die gemeinsamen Fehler lassen sich Verwandtschaften nachweisen.[8] Signifikante (Trenn-)Fehler (z.B. Textauslassungen) oder auch andere signifikante Versionen einzelner Textstellen, die andere Exemplare nicht aufweisen, können dagegen ein entscheidendes Indiz dafür sein, dass keine Verwandtschaft vorliegt.

Nun gibt es aber komplexere Fälle, in denen ein Kopist für die Kopie eines Textes mehrere Exemplare mit unterschiedlichen Versionen des Textes – und auch mit Fehlern – vorliegen hatte und in denen er, gerade wenn er gelehrt war, im Zweifelsfall einmal der Version aus der einen Vorlage, einmal der Version aus der anderen Vorlage den Vorzug gab. Darüber hinaus gibt es im Fall von *de interpretatione* zahlreiche Fälle, in denen der Kopist dann eine Vorlage im Sinne der Version, die er für besser hielt und die er in einer anderen Vorlage gefunden hat, korrigiert hat (z.B. durch Textrasuren).[9] Es entstehen so – im Fachjargon – kontaminierte Kopien,[10] die selbst wiederum die Vorlage für neue Kopien werden konnten. Je vielfältiger die Kontaminationen sind, desto schwieriger wird der Nachvollzug, welche Versionen eine Kopie vorliegen hatte.[11]

Die Schrift *de interpretatione*, die über Jahrhunderte für die Logikausbildung im Fokus des Interesses stand, zeichnet sich durch eine immense Kontamination aus. Diese ist so groß, dass einige Forscher meinen, dass die Erforschung der ganzen griechischsprachigen Überlieferung nicht möglich sei.[12] Konkret zu den logischen Schriften des Organon, zu denen auch *de interpretatione* gehört, ist im Fach sogar festgehalten worden, dass eine Erfassung der handschriftlichen Überlieferung nicht in einem Forscherleben erreichbar sei.[13] Dieser Meinung kann man nun begegnen, indem man die Basis der Betrachtung von Verwandtschaften auf Felder ausweitet, in denen weniger Kontaminationen vorliegen, in denen sich direkte Übernahmen evidenter nachweisen lassen und in denen quantitative Analysen in Verbindung mit qualitativen Analysen Fortschritte bei den hoch komplexen Forschungsgegenstand versprechen.

Ein solch erweiterter methodischer Ansatz sieht vor allem vor, Glossen-, Scholien-, Diagramm- und Kommentartraditionen zu erforschen. Denn wenn – wieder verkürzt dargestellt – ein Kopist z.B. eine Glosse oder ein erklärendes Diagramm aus einer ihm vorliegenden Handschrift übernommen hat, so unterliegt dies in der Regel keiner Kontamination. Verwandtschaftsverhältnisse können so einfacher nachvollzogen werden. Diese Traditionen sind also in Verbindung mit signifikanten Lesarten oder Fehlern zu berücksichtigen, wenn man in einer solchen Forschungsfrage weiterkommen möchte. Dieser Ansatz scheint zunächst paradox zu sein, denn er erhöht die Zahl dessen, was zu erforschen ist, obwohl schon auf der Grundlage des

[8] vgl. dazu etwa: Maas 1927.
[9] Um nur ein Beispiel aus der griechischsprachigen handschriftlichen Überlieferung von *de interpretatione* anzuführen, sei verwiesen auf die Rasur des ν im griechischsprachigen Wort ἀπόφανσις (‚Aussagesatz') zu ἀπόφασις (‚verneinter Aussagesatz') im Codex Wien Vindebonensis Phil. Gr. 300, f. 139v.
[10] vgl. dazu etwa West 1973.
[11] Am deutlichsten auf den Punkt gebracht hat dies für *de interpretatione* Montanari 1984, Bd. 1, hier z.B. S. 50-51.
[12] vgl. zu diesem Urteil ebd., Bd. 1, S. 49ff.
[13] Dies ist das Urteil von Reinsch 2001, S. 57-69.

herkömmlich betrachteten Materials die Frage in einem Forscherleben als nicht beantwortbar eingeschätzt wurde. So ist allein die Zahl der Glossen und Diagramme so groß, dass man ihrer nur schwer Herr werden kann, weil es extrem schwierig ist, den Überblick zu behalten. Ferner ist der aufzubringende Zeitaufwand für ein Entziffern der oft nur sehr schwer lesbaren Glossen und Scholien nicht zu unterschätzen. Andererseits kann gerade an diesem Punkt aufgezeigt werden, wie die Ausweitung der Betrachtung durch die Mittel einer digitalen Forschungsinfrastruktur und neue Praktiken der Zusammenarbeit, die diese ermöglicht, durchaus Sinn ergibt.

4. ‚Analoge' Materialien, Materialaufbereitungen und ihre Grenzen

Für jeden, der sich von der Faszination an der Erforschung komplexer Überlieferung bannen lässt, bietet sich, wenn er auf ‚analogem' Weg und mittels der Methode der Kollation fortschreitet, allerdings immer wieder die Ausgangssituation, fast von vorne anfangen zu müssen. Die Forscher begründen nämlich in ihren Publikationen die Unerforschbarkeit stets damit, dass die Kontaminationen so groß seien. Die Textkollationen aber, die die Basis für ein solches Urteil bilden, werden dabei, gerade weil das Urteil negativ ist und nicht etwas positiv nachgewiesen wird, größtenteils nicht veröffentlicht und sind nicht zugänglich.[14]

Unerlässlich für die Erforschung dieser Fragen sind im Feld der antiken griechischen und lateinischen Texte Handschriftenbeschreibungen, die sich in gedruckten Katalogen zum handschriftlichen Bestand einer Bibliothek finden. Diese beinhalten mindestens Datierungen oder Hypothesen zu einer solchen, die Übersicht über die in einem Codex enthaltenen Schriften und die Lagen- oder Folienzahl des Codex. In einigen – aber nicht in allen – Fällen finden sich auch Angaben zu der Provenienz einer Handschrift, in wenigeren Fällen Angaben zum Kopisten.[15] Bei Aristoteles sind wir für viele Handschriften in der glücklicheren Situation, dass es in den vergangenen fünf Jahrzehnten umfassende Handschriftenbeschreibungen der ca. 1000 bekannten Aristoteleshandschriften gegeben hat. Glücklich an dieser Situation ist, dass die vorliegenden Handschriftenbeschreibungen deutlich umfangreicher sind als die, die für andere Autoren verfügbar sind. Sie beinhalten etwa ausführliche und höchst gelehrte paläographische und kodikologische Beschreibungen, auf deren Basis auch für die Fälle, in denen es in der Handschrift selbst keine Vermerke gibt, fundierte Hypothesen zur Provenienz, Datierung oder auch zur Herkunft und Identität des Kopisten geäußert werden. Ebenso gibt es – allerdings nur in sehr wenigen Fällen – thesenhaft Angaben zu denkbaren Verwandtschaften unter Handschriften. In fast allen Fällen wird aber die konkrete Basis, die zu diesem Schluss führt, nicht angeführt (konkret die Ergebnisse von Textkollationen: z.B. welche konkreten gemeinsamen Fehler Handschriften aufweisen o.ä.).

Von den meisten der 1000 Handschriften selbst gibt es immerhin Microfilme, die im Aristoteles-Archiv an der Freien Universität Berlin liegen. Dazu gibt es einen wachsenden Bestand von

[14] Dies gilt für alle in Anm. 5 und 7 angeführten Titel. Eine gewisse Ausnahme bietet (thesenartig) Weidemann 2014. Weidemanns Angaben beschränken sich allerdings auf vergleichsweise wenige Fälle aus nur acht der 150 griechischsprachigen Handschriften. Auch in den textkritischen Apparaten der bisherigen Ausgaben beschränken sich die Angaben von unterschiedlichen Lesarten lediglich auf die berücksichtigten Handschriften. Die Anzahl dieser variiert je nach Textausgabe zwischen zwei und acht griechischsprachigen Exemplaren.

[15] vgl. für nur ein charakteristisches Beispiel: Mioni, Elpisio: Bibliothecae Divi Marci Venetiarum Codices Manuscripti Graeci, 5 Bde. Rom: Istituto poligrafico dello Stato, Libreria dello Stato 1960-1986.

Farbdigitalisaten, die entweder von Bibliotheken online gestellt werden,[16] oder von denen Kopien käuflich erworben werden können.

Das Problem der ‚analogen' Forschungen liegt nun darin, dass die Ergebnisse der Beschreibungen von Aristoteles bislang höchstens zu einem Drittel publiziert sind.[17] Da diese Publikation noch aus den 1970er Jahren stammt, ist auch nicht damit zu rechnen, dass die übrigen Ergebnisse noch in Druckform nachhaltig zugänglich werden. Es gibt nun erfreulicherweise an der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW) ein Projekt zu Aristoteleskommentatoren, innerhalb dessen zumindest die Handschriftenbeschreibungen in ihrer Ausführlichkeit Stück für Stück online gestellt werden.[18] Doch es ist zu vermuten, dass es noch eine ganze Weile dauert, bis die umfangreichen Handschriftenbeschreibungen einmal weitestgehend vollständig publiziert sein werden. Unbestreitbar ist aber schon jetzt der Vorteil, dass die Onlineversionen der Handschriftenbeschreibungen leicht und schnell verfügbar sind. Für die dargelegte Bearbeitung des Forschungsgegenstandes bedeutet dies, dass zum einen noch in einem nicht geringen Umfang auf Handschriftenbeschreibungen zurückgegriffen wird, und dass zum anderen die vorliegenden Daten zu den Handschriften durch eigene Forschungen und eigene Kollationen gerade mit Hinblick auf mögliche Verwandtschaftsverhältnisse noch ergänzt werden.

5. Forschungsdesiderate und Anforderungsprofil für eine digitale Infrastruktur

Angesichts dieser Ausgangslage ergeben sich die Forschungsdesiderate, denen sich eine eine digitale Forschungsinfrastruktur widmen kann, fast wie von selbst:

(1) Die Forschungsergebnisse, die erzielt wurden, konkret: die signifikanten Fehler oder auch Lesarten und auch Transkriptionen von Glossen, Scholien, Diagrammen und Kommentaren müssen nachhaltig abgelegt werden können, so dass die Ergebnisse nicht mit dem Ende eines Projektes oder eines Forscherlebens verschwinden – gerade weil diese konkrete Arbeit nicht in Form von gedruckten Publikationen erscheint, weil dort meist nämlich das Ergebnis am Ende nur negativ zusammengefasst und kein positiver Nachweis geführt wird.

(2) Für die Transparenz und Nachvollziehbarkeit der Kollationen und Transkriptionen ist es wünschenswert, diese unmittelbar in Verbindung mit dem Original vorliegen zu haben. Hierfür ist die Zusammenführung von digitalen Scans jeder einzelnen Handschriftenseite (Folium) in Form von Microfilmen oder Farbdigitalisaten an einem Ort unerlässlich. Die signifikanten Fehler und Lesarten sowie die Transkriptionen von Glossen usw. sollten mit eindeutiger Referenz auf das Digitalisat eines einzelnen Foliums vermerkt werden können.

[16] vgl. nur exemplarisch die stets wachsenden Ressourcen von digitalisierten und online zugänglichen Handschriften der Biblioteca Vaticana: https://digi.vatlib.it/ (Zugriffsdatum: 10.10.2018).

[17] vgl. Moraux, Paul u.a.: Aristoteles Graecus, Bd. 1. Alexandrien–London. Berlin, New York: De Gruyter 1976.

[18] vgl. https://cagb-db.bbaw.de/ (Zugriffsdatum: 27.08.2018). Darüber hinaus finden sich (ohne ausführliche Beschreibungen) einige Angaben zu den Metadaten der einzelnen Handschriften unter: http://pinakes.irht.cnrs.fr/notices/oeuvre/2973/ (Zugriffsdatum: 27.08.2018).

(3) Bei dem enormen Material und seiner Vielfältigkeit können quantitative Auswertungen den ersten wichtigen Schritt bilden, um Gruppen von Handschriften, die Gemeinsamkeiten aufweisen, zu identifizieren. In einem zweiten Schritt können genauere Verwandtschaftsverhältnisse durch feinere qualitative Analysen untersucht werden, die u.a. auch die Unterschiede evaluieren. Digitale Hilfen für eine quantitative Analyse können so höchst bedeutsam werden. Für eine quantitative Auswertung wäre es hilfreich, wenn die verschiedenen Parameter (z.B. gleiche Lesarten oder Fehler in Verbindung mit gleichen Glossen, Scholien, Diagrammen oder Kommentaren) miteinander in Verbindung gesetzt und verlässlich ausgewertet werden können. Oder anders formuliert: Die Handschriften sollten mit Blick auf gleiche Lesarten, Scholien, Glossen, Diagramme und Kommentare in Verbindung miteinander auf Gemeinsamkeiten durchsuchbar gemacht werden. In dem Fall, in dem Handschriften z.B. auf gemeinsame Fehler hin durchsuchbar sind, kann sich eine qualitative Auswertung (Fehler, aber auch z.B. signifikante Glossen, signifikante Diagramme) mit einer quantitativen Auswertung verbinden, z.B.: Teilen die Handschriften alle oder fast alle signifikanten Glossen und Fehler?

(4) Eine Verbindung mit ausgewählten Metadaten zu den Handschriften (v.a. Datierung, örtliche Provenienz), wie sie sich in analogen Handschriftenbeschreibungen finden, ist unverzichtbar für die inhaltliche Erschließung. Wenn etwa unter quantitativen wie qualitativen Gesichtspunkten Gemeinsamkeiten unter Handschriften gefunden werden, kann z.B. bei gleichem Ort und unterschiedlicher Zeit auf bestimmte Zentren oder Regionen geschlossen werden, in denen die Handschriften innerhalb eines bestimmten Zeitraums besonders oft kopiert wurden, so dass auf ein verstärktes Interesse an dieser Schrift in diesem Zeitraum geschlossen werden kann. Oder ein anderes Beispiel: Wenn unter quantitativen wie qualitativen Gesichtspunkten signifikante Gemeinsamkeiten gefunden werden, die Handschriften aber an geographisch verschiedenen Orten zu unterschiedlichen Zeiten lokalisiert werden können, können Rückschlüsse über den Verbreitungsweg und die Reisen von Manuskripten mit bestimmten Lesarten und Diagrammen, Glossen und Scholien innerhalb eines Zeitraums gewonnen werden.

6. Neue Möglichkeiten durch eine digitale Infrastruktur

Die Schaffung einer digitalen Infrastruktur kann in vielfältiger Weise helfen, diese Desiderate für Fortschritte bei der Lösung der Forschungsfrage aufzugreifen oder gar zu erfüllen. Im SFB 980 steht als zentrale infrastrukturelle Komponente ein Datenrepositorium zur Verfügung. In diesem Repositorium werden Daten inklusive ihrer beschreibenden Metadaten strukturiert verwaltet. Die Registrierung standardisierter Metadatenschemata garantiert dabei die Validität der Metadaten und unterstützt eine nachhaltige und interoperable Nutzung der abgelegten Forschungsergebnisse. Dabei geht es nicht nur um eine reine Datenablage. Vielmehr ermöglichen Import- und Exportschnittstellen die Anbindung diverser Dienste zur Ergänzung und Auswertung der Forschungsdaten – und somit eine direkte Integration des Repositoriums in die tägliche Arbeit der Forschenden. Die durch die digitale Infrastruktur ermöglichten zusätzlichen Forschungsmethoden und -verfahren sollen im Folgenden wenigstens kurz an dem Fallbeispiel aus dem SFB 980 skizziert werden:

(1) Es können mittlerweile problemlos digitale Scans von Microfilmen angefertigt werden. Falls ausreichend finanzielle Mittel verfügbar sind, können auch Farbdigitalisate meist einfach beschafft werden. Unser Datenrepositorium bietet die Möglichkeit, die Digitalisate aller Handschriftenseiten für Forschungszwecke zentral zu sammeln und den Forschenden dabei ortsunabhängig zugänglich zu machen. Die Sammlung besteht perspektivisch aus allen Folia der Handschriften, die die Schrift *de interpretatione* überliefern sowie Metadaten zu den Codices. Die gesamte griechischsprachige Überlieferung ist dann an einem virtuellen Ort gebündelt. Diese Form der Verfügbarkeit und einfachen Aufrufbarkeit besitzt für die Forschenden an diesem Gegenstand große Vorteile z.B. gegenüber dem herkömmlichen Lesen von Microfilmen oder gar gegenüber Bibliotheksreisen. Über ein solches Repositorium sind die Handschriften (nach Klärung lizenzrechtlicher Fragen) vor allem nachhaltig nutzbar und recherchierbar.

(2) An ein solches Repositorium können nun digitale Tools angeschlossen werden, die auch für mehrere Projekte mit jeweils eigenen Daten, Metadaten und Modellen adaptiert werden können. Im Falle des Projektes zu den *de interpretatione*-Handschriften sind dies z.B. eine automatische und halbautomatische Layoutanalyse sowie ein Annotationstool. Über das Annotationsprogramm können nun im vorliegenden Fall Glossen, Scholien, Diagramme oder auch Kommentare auf der digitalen Handschriftenseite als solche ausgezeichnet, transkribiert oder übersetzt werden. Werden etwa die Provenienz von Scholien und Kommentaren in Annotationen vermerkt (z.B. dass die Scholien Auszüge aus dem Text eines spätantiken Kommentars zu dem Werk sind, oder dass die Scholien von einem Lehrer der Philosophie aus der Zeit der Kopie der Handschrift stammen), so können Rückschlüsse auf Auslegungen oder auch didaktischen Verwendungen der Manuskripte gewonnen werden.

Durch Annotationen können die Forschungsgegenstände zudem unter Hinzuziehung projekt- oder fachspezifischer standardisierter Vokabulare (z.B. CodiKOS[19] und Tadirah) weiter beschrieben werden. Ebenso können die signifikanten Lesarten oder auch Fehler unmittelbar im Digitalisat als solche markiert und transkribiert werden. Qualitative Annotationen werden so zugleich auch für weitere Auswertungen zugänglich. Damit sind auch konkret in der digitalen Infrastruktur die Voraussetzungen dafür geschaffen, Gemeinsamkeiten oder Unterschiede der Handschriften zueinander quantitativ zu analysieren.

Die vielfältigen Vorteile, die diese Infrastruktur eröffnet, sind angesichts der dargestellten Problemlage nun unübersehbar: (i) Die zentralen Ergebnisse von Kollationen, die aufgrund der Einschätzungen, dass man die Überlieferungsgeschichte und -wege nicht erforschen könne, nicht näher veröffentlicht wurden, gehen jetzt nicht mehr verloren, sondern bleiben abgelegt und nachhaltig verfügbar. Es ist besonders darauf hinzuweisen, dass die Arbeit mit digitalen Tools die herkömmlichen Methoden und Forschungspraktiken keinesfalls in Zweifel zieht, sondern diese vielmehr ergänzt. Sie verhelfen dazu, die mit diesen Methoden erzielten Resultate

[19] Ontologie zur Beschreibung kodikologischer Termini, das im Rahmen des Projektes eCodicology: http://www.e-codicology.org (Zugriffsdatum: 07.09.2018) erstellt wurde. Die persistente Veröffentlichung im Rahmen des SFB 980 erfolgt in Kürze. Im beschriebenen Vorhaben wird ein Auszug des Vokabulars genutzt.

nachhaltig verfügbar zu machen und in Kombination mit weiteren digitalisierten Forschungsdaten (Daten zu den Handschriften und transkribierten und teils beschriebenen Glossen, Scholien, Diagrammen und Kommentaren) auszuwerten, um bislang nur schwer erkennbare Zusammenhänge leichter aufzuspüren. (ii) Auf der Basis transkribierter Glossen, Scholien, Diagramme und Kommentare sowie vermerkter signifikanter Fehler können, indem z.B. (qualitativ) signifikante Fehler als solche ausgewiesen werden, Handschriften oder Handschriftengruppen, die quantitativ und qualitativ wichtige Gemeinsamkeiten aufweisen, leicht identifiziert werden. Wenn z.B. vergleichsweise wenige wichtige qualitative Gemeinsamkeiten identifiziert werden, die Handschriften ansonsten quantitativ zahlreiche Verschiedenheiten in der Glossierung, bei den Diagrammen aufweisen usw., können die für den Nachvollzug der Überlieferungsgeschichte wichtigen Kreuzungspunkte von Handschriften mit verschiedenen Lesarten gefunden werden, an denen Kontaminationen entstanden.[20] (iii) Es gibt die Möglichkeiten eines gemeinsamen und interoperablen Arbeitens von unterschiedlichen Standorten und auch Institutionen aus an demselben Material. Es ergeben sich die Möglichkeiten für Praktiken des Forschens, die sich höchst fruchtbar erweisen, weil die Ergebnisse der verschiedenen Forscher von dem jeweils anderen mitverwendet werden können. (iv) Ergebnisse aus ‚analog' publizierten Handschriftenbeschreibungen können in XML (TEI) übertragen werden und so unmittelbar in die automatischen Analysen und Auswertungen einbezogen werden. Die Forschungsdaten sind damit nicht mehr nur leicht online verfügbar, sondern können gleich produktiv in die Auswertungen durch digitale Tools integriert werden.

Die beschriebene digitale Infrastruktur mit den nun ermöglichten Praktiken des interoperablen Forschens und der Verbindung von quantitativer und qualitativer Auswertung vermag so perspektivisch auch die aufgezeigte Erweiterung des zu bearbeitenden Materials über die Integration von Glossen, Scholien, Diagrammen und Kommentaren in die Betrachtungen aufzufangen. Gerade die Möglichkeit, dass man letztere nun berücksichtigen und auswerten kann, verspricht große Fortschritte in der bestehenden Forschungsfrage.

7. Vom Fallbeispiel zum Ausblick für weitere Projekte im SFB 980

Die an einem Fallbeispiel aufgezeigten Vorteile sind auch für weitere Projekte z.B. innerhalb eines Sonderforschungsbereichs wie dem SFB 980 gegeben. Für alle beteiligten Projekte bietet das Datenrepositorium zunächst die Möglichkeit, umfangreiche Materialien bzw. Forschungsgegenstände schnell und leicht zugänglich verfügbar zu machen. Ausgehend von konkreten Forschungsfragen können digitalisierte Sammlungen mit Daten und Metadaten angelegt werden.

Die Erfahrung zeigt, dass auch digitale Tools, die ausgehend von einer konkreten Forschungsfrage entwickelt wurden, bei ausreichend generischer Konzeption von weiteren Projekten nachgenutzt werden können. Ein Projekt, das sich z.B. mit Fragen zu Sprachlernhandbüchern beschäftigt, kann die digitalisierten Handbücher ebenfalls im Repositorium ablegen. Bereits im Projekt vorliegende Annotationen können, einmal in das *Web Annotation Data Model*

[20] vgl. zu konkreten Beispielen exemplarisch: Krewet, Michael / Hegel, Philipp: Diagramme in Bewegung: Scholien und Glossen zu *De interpretatione*, erscheint in: Canan Hastik / Philipp Hegel (Hg.): Bilddaten in den digitalen Geisteswissenschaften, Wiesbaden 2019 (Episteme).

übertragen, in die digitale Infrastruktur importiert werden. Mit dem mit einem projektspezifischen Vokabular angepassten Annotationsprogramm ist es möglich, die Digitalisate ebenfalls nach bestimmten Kategorien zu annotieren. Jede Person, die in diesem Projekt tätig ist, kann alle oder auch einzelne Handbücher nun bezüglich der Kategorien (z.B. eine bestimmte Person im Kontext verschiedener idiomatischer Wendungen) durchgehen und entsprechend annotieren. Ferner werden die Metadaten (z.B. zur Datierung oder auch der Provenienz oder dem Ort der Verwendung) zu den Drucken ebenfalls in eine XML-Struktur (TEI) überführt. Die digitalisierten Drucke können damit später gleichermaßen nach Gemeinsamkeiten durchsucht werden, woraus z.B. konkrete Rückschlüsse gewonnen werden, in welcher Form oder an welchen Orten oder zu welchen Zeiten eine Sprache gelernt wurde, ob die Grammatik eine Bedeutung hatte, ob eher an Fallbeispielen und über konkrete Anwendungssituationen gelernt wurde, ob Beispiele aus anderen Lehrbüchern übernommen wurden usw. Die Infrastruktur mit den dargestellten Tools ermöglicht in einem solchen Fall, auch Forschungsfelder mit einer umfangreichen Materialbasis in einer absehbaren Zeit erschließen zu können. Das Repositorium und daran angeschlossene Werkzeuge stehen in dieser Art auch für andere Projekte im Sonderforschungsbereich zur Verfügung.

Denkt man an solche Nutzungen, die in Forschungsverbünden wie Sonderforschungsbereichen über das einzelne Projekt hinausgehen, dann können sich verschiedene technische, fachwissenschaftliche und institutionelle Fragen stellen. So liegt vielleicht der Gedanke an standardisierte Verfahren und kontrollierte Vokabulare als Grundlage für die projektübergreifende Nutzung nahe. Taxonomien und Normdaten stecken jedoch in einigen Disziplinen noch in den Anfängen. Man kann sie weder bei der Arbeit innerhalb des Fachs noch bei der Errichtung einer Infrastruktur als gegeben betrachten. So kann beispielsweise die Eigenheit einer Anmerkung in einem Codex Indiz für eine historisch relevante Aneignung in einer spezifischen Situation sein, die nur rudimentär mit Hilfe bestehender Normdaten zu Orten und Personen erfasst werden kann. Dennoch stellen diese manchmal sehr speziellen Befunde zu Form und Inhalt der Anmerkungen als Anzeichen für Schulen und Traditionen oder andere ‚epistemische Konstellationen' Befunde dar, die auch als Forschungsdaten erhalten bleiben sollen.

Vielfach reicht die „Ausdruckskraft" gegebener Vokabulare und Taxonomien nicht hin, um dem spezifischen Erkenntnisinteresse vollauf gerecht zu werden.[21] Oftmals kann diesem Problem begegnet werden, indem die Möglichkeit zur Erweiterung bestehender oder zur Entwicklung eigener, projektspezifischer Ontologien geboten wird. Gelegentlich muss aber auch dem Anteil an Forschungsdaten und -ergebnissen, der über die verwendeten Klassifikationen hinausreichen, Platz, und das heißt in der Regel ‚Freitext', eingeräumt werden. Während bei Diagrammtypen eine für das Forschungsvorhaben passende Klassifikation vorgenommen werden konnte, ist dies bei Transkriptionen und Übersetzungen etwa zu einzelnen Diagrammen oder Glossen nicht einfach zu erreichen. Sie dienen zwar einem textkritischen Interesse, variieren aber bereits innerhalb dieser Überlieferung und lassen sich nicht problemlos und sinnvoll auf andere Werke und Forschungsvorhaben übertragen.

[21] So auch Bauman, Syd: Interchange vs. Interoperability. In: Proc. Balisage 7 (2011). http://www.balisage.net/Proceedings/vol7/print/Bauman01/BalisageVol7-Bauman01.html (Zugriffsdatum: 19.11.2018), unpaginiert, mit dem Blick auf Auszeichnungssprachen: „since interoperationality (equality) is often bought at the expense of expressivity (liberty), interoperability is the wrong goal for scholarly humanities text encoding. Practitioners of this activity should aim for blind interchange."

In einigen Fällen sind Interoperabilität oder das von Syd Bauman befürwortete Ziel des „blinden Austauschs“ durch Verwendung von Standards und Dokumentation über das einzelne Anliegen hinaus möglich und zweckmäßig.[22] In anderen Fällen ist aber, wie in einem Sonderforschungsbereich etwa, auch ein zusätzlicher, „verhandelter Austausch“ zwischen den Akteuren möglich, um zu bestimmen, ob und wie eine breitere Nutzung erreicht werden kann.[23] Ein solch verhandelter Austausch als einem Sprechen nicht nur über Daten und Gegenstände, sondern auch über ihre Formen und Modelle ist dabei in der interdisziplinären Zusammenarbeit von Geisteswissenschaften und Informatik gerade nicht nur „geschwätzig“.

Digitale Modellierungen geisteswissenschaftlicher Gegenstände können nämlich bekanntermaßen auf Forschungsfragen rückwirken und den Erkenntnisprozess begleiten.[24] In diesen Modellen werden zum Beispiel relevante Eigenschaften der Gegenstände, ihr Charakter, die Abhängigkeiten und möglichen Werte dieser Eigenschaften sowie die Relationen zwischen Gegenständen festgehalten. ‚Remodellierungen‘ werden je nach Erkenntnisinteresse und technischen Werkzeugen erforderlich.

Datenmodelle können darüber hinaus jedoch auch das Potential zu einer Standardisierung und einem „blinden Austausch“ und Interoperabilität in sich bergen. Sie sind im Entstehen aber oft an kontinuierliche Gespräche zwischen Vertretern verschiedener geisteswissenschaftlicher Disziplinen einerseits und Experten für die Umsetzung in digitaler Infrastruktur andererseits gebunden. Die gemeinsame Arbeit am Modell kann Anlass sein, eine interdisziplinäre Sprache zu suchen.

In technischen Infrastrukturen für Sonderforschungsbereiche potenzieren sich diese Erfordernisse, weil nicht nur die Interessen eines einzelnen Vorhabens zu berücksichtigen sind. Diese Infrastrukturen gehen über das einzelne Projekt hinaus, haben aber nicht den Anspruch, allen Bedürfnissen einer Fachgemeinschaft oder gleich mehrerer Fachgemeinschaften nachzukommen. Sonderforschungsbereiche sind auch ein Ort, um zu erkunden, wie Erweiterungen von Infrastrukturen aussehen können, um Projekte zu unterstützten, von denen man zum jetzigen Zeitpunkt noch wenig oder gar nichts weiß. Mit ihrer Vielfalt an Fächern und Vorhaben stellen Sonderforschungsbereiche und ihre Infrastrukturprojekte eine besondere, nicht immer einfache, aber vielleicht doch nahezu ideale Grundlage dar, um Aushandlungsprozesse zwischen Fachwissenschaft und Technik über einen längeren Zeitraum zu gestalten.

---

[22] vgl. ebd.: „Human intervention, but not direct communication, is required.“ In seiner Zusammenfassung präferiert er dieses Verfahren für die von ihm behandelten Fälle: „Let us hereby decree our common goal of supporting blind interchange over mindless interoperability and chatty negotiated interchange.“

[23] gl. ebd.: „Both human communication and human intervention are required.“

[24] Es ist oftmals betont worden, dass sich die Modellierung von Daten und Forschungsfragen beeinflussen. Vgl. etwa Ciula, Adrianna / Eide, Øyvind: Modelling in digital humanities: Signs in Context. In: Digital Scholarship in the Humanities 32, Suppl. 1 (2016), S. 33-46. DOI: 10.1093/llc/fqw045 (Zugriffsdatum: 19.11.2018), hier S. 34: „modelling is a process by which researchers make and manipulate external representations […] to make sense of the conceptual objects and phenomena they study.“ McCarty, Willard: Modeling: A Study in Words and Meaning. In: Schreibman, Susan / Siemens, Ray / Unsworth, John (Hg.): A Companion to Digital Humanities, Oxford: Blackwell 2004. http://www.digitalhumanities.org/companion (Zugriffsdatum: 19.11.2018), unpaginiert, schreibt zum diesem Verhältnis: „modeling *of* something readily turns into modeling *for* better or more detailed knowledge of it; similarly, the knowledge gained from realizing a model *for* something feeds or can feed into an improved version.“

Literaturverzeichnis

Bauman, Syd: Interchange vs. Interoperability. In: Proc. Balisage 7 (2011). http://www.balisage.net/Proceedings/vol7/print/Bauman01/BalisageVol7-Bauman01.html (Zugriffsdatum: 19.11.2018).

Ciula, Adrianna / Eide, Øyvind: Modelling in digital humanities: Signs in Context. In: Digital Scholarship in the Humanities 32, Suppl. 1 (2016), S. 33-46. DOI: 10.1093/llc/fqw045 (Zugriffsdatum: 19.11.2018).

Crisci, Edoardo / Degni, Paola: La Scrittura Greca dall' Antichità All' Epoca della Stampa. Rom: Carocci 2011.

Maas, Paul: Textkritik. Leipzig, Berlin: Teubner 1927.

McCarty, Willard: Modeling: A Study in Words and Meaning. In: Schreibman, Susan / Siemens, Ray / Unsworth, John (Hg.): A Companion to Digital Humanities, Oxford: Blackwell 2004. http://www.digitalhumanities.org/companion (Zugriffsdatum: 19.11.2018)

Minio Paluello, Lorenzo (Hg.): Aristotelis Categoriae et Liber de interpretatione. Oxford: Clarendon 1949.

Mioni, Elpisio: Bibliothecae Divi Marci Venetiarum Codices Manuscripti Graeci, 5 Bde. Rom: Istituto poligrafico dello Stato, Libreria dello Stato 1960-1986.

Montanari, Elio: La sezione linguistica del *Peri Hermeneias* di Aristotele, 2 Bde. (Studi e testi 5) Florenz Università degli studi die Firenze, Dipartimento di scienze dell'antichità "Giorgio Pasquali" 1984.

Moraux, Paul u.a.: Aristoteles Graecus, Bd. 1. Alexandrien–London. Berlin, New York: De Gruyter 1976.

Reinsch, Diether Roderich: Fragmente einer Organon-Handschrift des zehnten Jahrhunderts aus dem Katharinenkloster auf dem Berg Sinai. In: Philologus 145,1 (2001), S. 57-69.

Weidemann, Hermann (Hg.): Aristoteles, de interpretatione. Berlin, New York: De Gruyter 2014.

West, Martin L.: Textual Criticism and Editorial Technique: applicable to Greek and Latin texts. Stuttgart: Teubner 1973.

„Das ist nicht ganz trivial …“. Die Anpassung gewachsener Projektstrukturen an moderne IT-Standards am Beispiel des *Handschriftencensus*

Bernhard Runzheimer (Marburg)

Während die Verbindung von Informatik und Geisteswissenschaften in der Regel bereits eine kommunikative Herausforderung – sowohl auf struktureller als auch auf menschlicher Ebene – darstellt, sind es vor allem die Anpassungen der verfügbaren digitalen Infrastrukturen an gängige Standards, die einen nicht unerheblichen Anteil an Arbeit und Zeit einfordern. Hinzu kommt, dass speziell in geisteswissenschaftlichen Projekten mit hoher Fachspezialisierung der inhaltliche Zugang für Informatiker deutlich schwieriger ist, da die Gefahr besteht, dass beide Seiten aufgrund fehlender Schnittmengenkompetenz aneinander vorbeireden und das Projekt dadurch Schaden nimmt. Besonders Projekte, deren Umfang sich durch konstante Anforderungsanpassungen über mehrere Jahre exponentiell erweitert haben, sehen sich irgendwann mit dem Problem des digitalen ‚breaking point‘ konfrontiert, an dem es sinnvoller wäre, die überholten Infrastrukturen einer kompletten Neumodellierung zu unterziehen anstatt diese bis zur Unbenutzbarkeit zu verschlimmbessern.

Der *Handschriftencensus* am Institut für deutsche Philologie des Mittelalters an der Philipps-Universität Marburg hat sich zum Ziel gesetzt, das gesamte deutschsprachige Handschriftenerbe des Mittelalters systematisch zu erfassen und wird seit 2017 für 20 Jahre von der Mainzer Akademie der Wissenschaften und der Literatur gefördert. Die Datenbank des HSC entstand seinerzeit als Projekt einer Hilfskraft und wurde einem ständig wachsenden Bedarf angepasst und lange Zeit ehrenamtlich betreut. Mittlerweile umfasst die Datenbank nicht nur die Handschriften aus den DFG-geförderten „Marburger Repertorien deutschsprachiger Handschriften des 13. und 14. Jahrhunderts“ sondern auch die ähnlich strukturierten Beschreibungen der deutschsprachigen Handschriften bis 1200 im „Paderborner Repertorium der deutschsprachigen Textüberlieferung des 8. bis 12. Jahrhunderts“. Derzeit sind im HSC rund 24.000 Textzeugen meist rudimentär erfasst, davon aber bisher nur rund 5.000 Bucheinheiten in der intendierten Tiefe erschlossen: Für das Gros der spätmittelalterlichen Textzeugen besteht ein Datensatz bislang nur aus minimalen Bestandsdaten (Signatur) und einem Literaturhinweis

Als Akademie-gefördertes Projekt steht der *Handschriftencensus* nun vor der großen Aufgabe, die langjährig gewachsenen und technisch überholten Strukturen an die Erfordernisse moderner IT-Standards anzupassen. Dazu gehören vor allem die grundlegende Überarbeitung des Datenmodells und der Suchfunktion, die Bereitstellung von Schnittstellen für Datenim- und export, ein effektives Normdatenmanagement sowie die Aktualisierung der Webseitenarchitektur.

Dieser Beitrag möchte einen Überblick über das Projekt, die historische Entwicklung und die Veränderung der bestehenden Infrastrukturen geben, unter besonderer Berücksichtigung der dabei entstehenden Probleme und Herausforderungen.

1. Interdisziplinäre Kommunikation als primäre Hürde

> [T]he one who writes the code for the program. (or: The hidden master of the database. The magician.)[1]

Dieser Beitrag könnte mit der Floskel beginnen, dass die Geschichte der Zusammenarbeit von Natur- und Geisteswissenschaftlern voller Missverständnisse ist. Obgleich der Wahrheitsgehalt dieser Aussage im folgenden Beitrag nicht vollends entkräftet werden kann, sollen doch einige der damit einhergehenden Aspekte – speziell deren Auswirkungen in der interdisziplinären Projektarbeit – einer genaueren Betrachtung unterzogen werden. Viele Missverständnisse resultieren dabei nicht nur aus einer fehlgeschlagenen, sondern vielmehr aus einer nicht stattgefundenen Kommunikation. Die Vorurteile, die sowohl die Natur- als auch die Geisteswissenschaftler voneinander haben und nur zu gerne perpetuieren, halten sich hartnäckig: Auf der einen Seite dominiert das Bild der verkopften Dickbrettbohrer, deren endlose und redundante Monologe sich primär um die Legitimierung der eigenen Disziplin drehen, bevor man sich dann nach Jahren der vergeblichen Forschungsarbeit doch der Taxifahrerei ergibt oder dann ‚irgendwas mit Medien' in einer halben Stelle ‚im kulturellen Bereich' macht. Demgegenüber stehen die zumeist fülligeren und blassen Nerds mit Brille und Heavy-Metal-T-Shirt, die untereinander binär kommunizieren, sich von Energydrinks und Fertigpizza ernähren und ihr Dasein in einem dunklen Keller fristen, bis ihr dort gegründetes Startup hoffentlich irgendwann von Google aufgekauft wird.

Abseits dieser ermüdenden Klischees bleibt dennoch festzuhalten, dass die wechselseitige Akzeptanz und Zusammenarbeit nicht nur stark von der individuellen Kommunikationsfähigkeit abhängt, sondern auch von der jeweiligen Kompromissbereitschaft. Die Durchführung größerer Projekte erfordert auf der Führungsebene nicht nur ein hohes Maß an Organisationsfähigkeit, sondern auch entsprechende kommunikative Fähigkeiten, um die einzelnen Mitarbeiter zu begeistern und anzuleiten. Dass fehlgeschlagene Kommunikation als einer der häufigsten Gründe für gescheiterte Projekte genannt wird, ist unbestritten und wird medial oft rekontextualisiert:

[1] Zitat eines geisteswissenschaftlichen Projektleiters, der einem Kollegen per Mail in eigenen Worten schildert, was er unter ‚Programmierer' versteht.

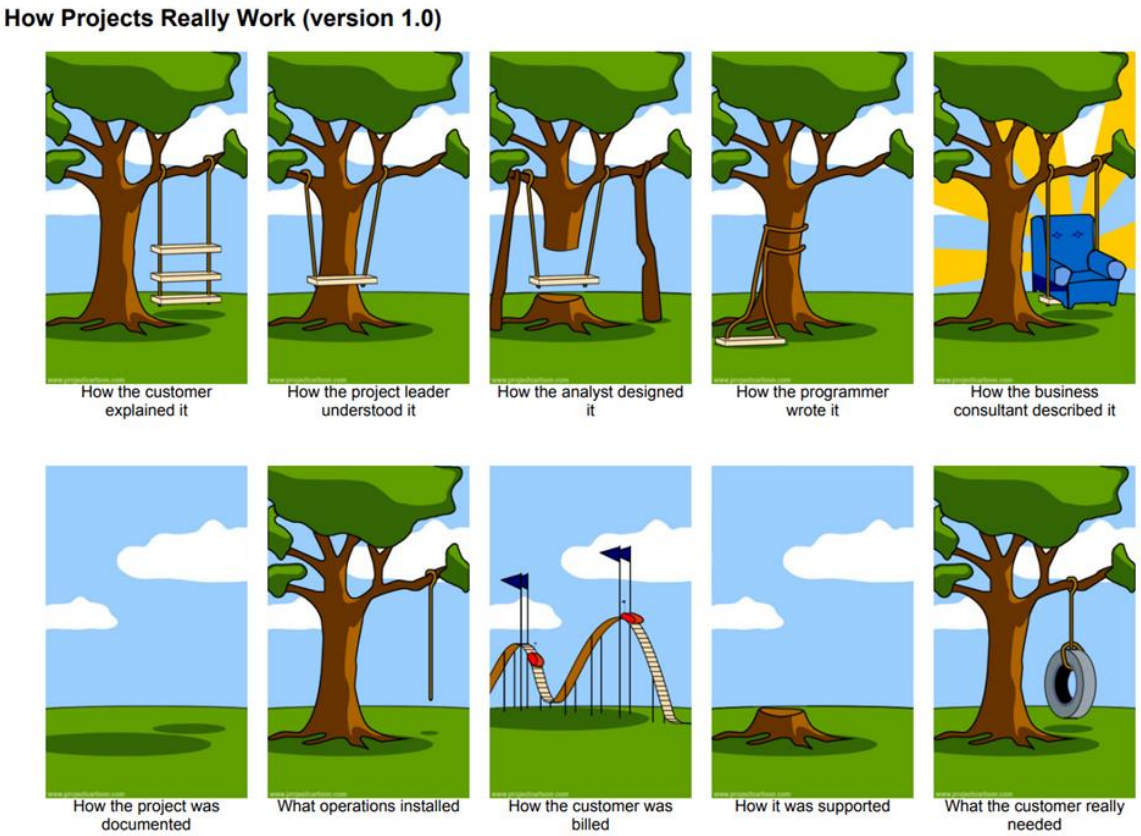

Abb.1: Kommunikative Missverständnisse in der Projektarbeit

Für den weiteren Verlauf dieses Beitrags sind lediglich die Einzelbilder 1, 4 und 10 relevant, da sie exemplarisch für die kommunikative Struktur in Digital-Humanities-Projekten stehen und das Verständnis der folgenden Projekthistorie und -entwicklung erleichtern.

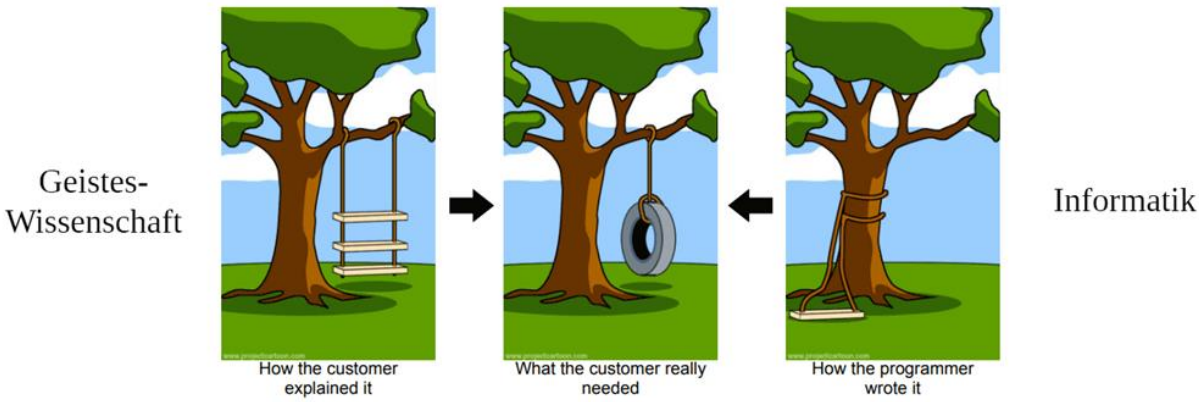

Abb.2: Exemplarische Projektvorstellungen von Geisteswissenschaftlern (links) und Informatikern (rechts) gegenüber dem eigentlichen Desiderat

2. Projekthistorie

Im Jahr 2006 gründete ein Verbund von Forschern aus Deutschland, Österreich und der Schweiz die AG *Handschriftencensus*, mit dem Ziel einer systematischen Erfassung aller deutschsprachiger Handschriften des Mittelalters. Das zu bedienende Desiderat entstand aus der bis dahin eher lückenhaft-heterogenen und in mehreren DFG-Projekten verteilten Handschriftendokumentation spezifischer Jahrhunderte, was für die Forschungscommunity nur bedingt produktiv nutzbar war.

Die historischen Anfänge gehen zurück auf das DFG-Projekt „Marburger Repertorium der deutschsprachigen Handschriften des 13. Jahrhunderts“, was am Marburger Institut für Deutsche Philologie des Mittelalters in den Jahren 1990-1999 durchgeführt wurde. Ursprünglich als Druckkatalog geplant, sollten darin der überschaubare Bestand aller dokumentierten mittelalterlichen Handschriften des 13. Jahrhunderts in einem Nachschlagewerk gebündelt werden, was jedoch letztlich im Resultat nicht realisiert wurde. In den darauf folgenden Jahren wurden zwei weitere Projekte angestoßen, die das Projekt MR13 ergänzten bzw. erweiterten: Während das „Marburger Repertorium der Freidank-Überlieferung“[2] (1998-2006) nur einen sehr spezifischen Teilaspekt umfasste, führte das „Marburger Repertorium der deutschsprachigen Handschriften des 14. Jahrhunderts“ (2004-2009) die bereits begonnene Arbeit fort und bündelte diese in einem digitalen Archiv, das die gesammelten Daten – seinerzeit noch als reine HTML-Variante – kostenfrei im Internet zur Verfügung stellte. In einer ab 2005 realisierten Datenbank wurden in der Folge die Daten von bis dahin drei DFG-Projekten von einer am Lehrstuhl des damaligen Projektleiters angestellten Hilfskraft archiviert, die sowohl Informatik-Kenntnisse als auch einen projektbezogenen geisteswissenschaftlichen Hintergrund in Personalunion vorweisen konnte. Diese Situation erwies sich als Glücksfall und erleichterte nicht nur den Arbeitsalltag der beteiligten Wissenschaftler enorm, sondern weckte auch Begehrlichkeiten: Warum sollte es nicht möglich sein, in dieser Datenbank auch Handschriften einzutragen, die sich außerhalb der Projektspezifikationen bewegen, aber dennoch wichtig für die Forschungsgemeinschaft sind?

Nach Gründung der AG Handschriftencensus war es das vorrangige Ziel, einen Überblick über die vorhandenen Textzeugen des deutschen Mittelalters und deren Spezifikationen und Verbleib zu bekommen und diese Informationen in der gemeinsam genutzten Datenbank abzulegen. Zu diesem Zweck versammelten sich in der AG Spezialisten aus unterschiedlichen Ländern, die die fehlenden Informationen in ehrenamtlicher Arbeit recherchierten und ergänzten. Im Jahr 2007 wurde zudem das verwandte Projekt „Paderborner Repertorium der deutschsprachigen Textüberlieferung des 8. bis 12. Jahrhunderts“ in die Datenbank eingegliedert, um die Chronologie zu vervollständigen. Dadurch gab es in dieser Zeit zwar einen exponentiellen Anstieg des Datenbestands, aber auch neue Hürden, die es zu überwinden galt: Während die nunmehr vier Projekte größtenteils dasselbe Datenhaltungsformat nutzten, mussten trotzdem spezifische Zusatzfelder angelegt werden, um den einzelnen Projektspezifikationen gerecht zu werden.

[2] Das „Marburger Repertorium der Freidank-Überlieferung“ verzeichnet in Form eines beschreibenden Katalogs die deutsche und lateinische Überlieferung der Sprüche Freidanks in Handschriften und Drucken bis zur letzten bisher nachgewiesenen Druck-Ausgabe von 1583 sowie in Auswahl die Überlieferung in Inschriften bis ins 17. Jahrhundert.

Die seit 2005 im Internet zugängliche Plattform umfasste daher vier verschiedene Projekte,[3] wovon der *Handschriftencensus* nach Ablauf aller Projektzeiten die übergeordnete Instanz darstellte, da er auf den bereits vorhandenen Daten aufbaute und diese durch zusätzliche Handschriften und Informationen ergänzte. Innerhalb der Forschungsgemeinschaft erwies sich der offen zugängliche Datenbestand als Glücksfall und wurde aufgrund der sorgfältigen und kompetenten Redaktion trotz des heterogenen Datenbestands eine häufig zitierte Quelle in der wissenschaftlichen Arbeit. Bereits nach kurzer Zeit meldeten sich viele Forscher, die die zum großen Teil noch unvollständigen Daten durch zusätzliche Hinweise und Korrekturen ergänzten, sodass ein interaktives Mitteilungsfeld eingerichtet wurde, um die Fülle an Informationen effektiv bewältigen zu können. Dies erwies sich aufgrund der Qualität der Mitteilungen zwar als wissenschaftlich äußerst wertvoll, führte jedoch aufgrund der Quantität der Einreichungen[4] zu einem Informationsstau, den die ehrenamtlichen Mitarbeiter in der ihnen zur Verfügung stehenden Zeit nicht mehr bewältigen konnten. Neben anderen Gründen musste das Mitteilungsfeld im Jahr 2015 deaktiviert werden.

Eine strukturelle Professionalisierung und Nachwuchsförderung wurde durch zwei DFG-Anträge angestrebt, die jedoch beide negativ beschieden wurden. In der Folge wurde der Projektantrag erweitert und bei der Mainzer Akademie der Wissenschaften als Langzeitprojekt vorgelegt, was im Jahr 2017 bestätigt wurde. Ursprünglich als ehrenamtlich betreute Bestandsübersicht initiiert, hat sich der *Handschriftencensus* somit im Laufe der vergangenen Jahre zu einer festen Größe in der mediävistischen Forschungslandschaft entwickelt, der nun bis ins Jahr 2036 als Akademie-Projekt fortgeführt wird.

3. Technische Struktur

3.1 Ist-Zustand

Der *Handschriftencensus* betreibt eine Datenbank und eine Online-Präsenz, auf der die Daten unter der URL www.handschriftencensus.de eingesehen werden können. Die Online-Präsenz dient den Projektmitarbeitern im eingeloggten Zustand gleichzeitig als Verwaltungs- und Eingabeoberfläche. Sämtliche Daten liegen auf einem Server des Hochschulrechenzentrums der Philipps-Universität Marburg und werden mithilfe einer Debian-Distribution (Version 6.0.10) und GitHub verwaltet. Die Datenbank verwendet MySQL 5.5.60. Für die Online-Präsenz wird PHP 5.3.3.7 und das CakePHP-Framework in der Version 1.2 genutzt. Legt man die verwendeten Versionsnummern zugrunde, befinden sich sämtliche Software-Komponenten auf dem Stand von 2010.

3.2 Soll-Zustand

Ausgehend von der im zweiten Kapitel beschriebenen Historie hat die Konsolidierung und Revision des Datenmodells aus mehreren Gründen höchste Projektpriorität. Durch die Kumulation mehrerer Einzelprojekte in einer Datenbank ergeben sich naturgemäß Strukturen, die nicht den gängigen IT-Standards folgen, sondern aus Notwendigkeiten heraus implementiert wurden. Viele Datenfelder wurden nachträglich hinzugefügt, um punktuellen Anforderungen

[3] Den *Handschriftencensus* (HSC), die beiden unter „MR13/14“ zusammengefassten Marburger Repertorien, das Paderborner Repertorium „PR08/12“ sowie das Freidank-Repertorium „MRFD“.
[4] Es handelt sich um knapp 20.000 Mitteilungen von über 450 Wissenschaftlern aus aller Welt.

eines ehrenamtlich geführten Projektes gerecht zu werden, sind aber in der derzeitigen Form nicht zukunftsfähig.

Daraus ergibt sich sowohl die Notwendigkeit einer Datenstandardisierung des bereits vorhandenen Materials als auch die Notwendigkeit für ein allgemeines Datenformat, das auch zukünftigen Änderungen standhält und dazu in der Lage ist, problemlos neue Datenmodelle zu implementieren, ohne das Grundmodell strukturell zu schwächen. Mit dieser soliden Basis können im Optimalfall nicht nur sämtliche Aspekte mittelalterlicher Handschriften digitalisiert, sondern diese im Anschluss durch Austausch-Schnittstellen in einem gängigen Standardformat der Öffentlichkeit zur Verfügung gestellt werden.

Dies schafft die Grundlage für eine umfassende Vernetzung mit bereits bestehenden Projekten, deren Daten dann wechselseitig genutzt werden können, um redundante Datenerfassung weitestgehend zu reduzieren. Dabei wird auch die Zusammenarbeit mit Normdatenstellen und die Nutzung von Union Resource Identifiern wie z. B. der Deutschen Nationalbibliothek, VIAF, DOI, Orcid, etc. angestrebt. Durch die Nutzung und Implementation offener Schnittstellen soll der Datenbestand des *HSC* sukzessive und automatisiert aufgefüllt werden, was bisher noch ausschließlich mittels manueller Eingabe der einzelnen Bearbeiter geschieht.

Durch die Kumulation mehrerer Datenquellen soll der *HSC* nicht nur seinen Stellenwert für die Handschriften-Community unterstreichen, sondern sich darüber hinaus auch als Informationsportal bzw. Schnittstelle für die breite Öffentlichkeit etablieren. Dazu ist es erforderlich, dass die Online-Präsenz sowohl technisch als auch grafisch überholt wird, um den gängigen Web-Standards zu entsprechen. Auf technischer Seite gehören dazu u. a. eine tragfähige Suchfunktion, ein modernes und mehrsprachiges Interface sowie eine Optimierung für die Anzeige auf mobilen Endgeräten.

4. Probleme

Aus der Projekthistorie des *HSC* ergeben sich einige typische IT-Probleme, die bei solchen „historisch gewachsenen“ Projekten eine Überarbeitung erschweren:

> Alte Systeme werden nach den gestiegenen bzw. veränderten Geschäftsanforderungen durch neue Systeme erweitert bzw. ergänzt – es entsteht eine ‚historisch gewachsene‘ Systemlandschaft mit einer hohen Anzahl von Schnittstellen und Systemen unterschiedlicher Herkunft, Funktionsweisen und unterschiedlichen Alters. Geschäftsprozesse sind dabei oft ‚hart verdrahtet‘ und tief im Code begraben – Änderungen sind nur schwer und teuer umzusetzen.[5]

Aus dieser Grundproblematik, die vor allem die von mehreren Projekten gemeinsam genutzte Datenbank betrifft, lassen sich alle weiteren Probleme direkt ableiten:

4.1 Software

Der Grundsatz ‚*never change a running system*‘ hat in IT-Projekten nur begrenzte Gültigkeit – vor allem, wenn es um Sicherheitslücken und den Schutz vor potentiellen Angreifern geht. In

[5] Göldner, Axel / Rüter, Andreas / Schröder, Jürgen (Hg.): IT-Governance in der Praxis. Berlin/Heidelberg: Springer 2010, S. 61.

dieser Hinsicht müssen Systemadministratoren vorausschauend tätig werden, um die verwalteten Systeme nicht erst dann abzusichern, nachdem etwas passiert ist. Im Falle der Datenbank und der Online-Präsenz des *HSC* wurde mit der Verwendung von CakePHP zwar ein stabiles Framework als Grundlage verwendet, jedoch im Laufe der Jahre versäumt, dieses regelmäßig auf die neueste Version zu aktualisieren. Zwischen der verwendeten Version 1.2 und der derzeit aktuellen Version 3.6 liegen zwei große Versionssprünge, die neben diversen Sicherheitsupdates auch größere Änderungen in der Architektur beinhalten, wodurch eine Aktualisierung in diesem Zustand vermutlich mehr Aufwand bedeuten würde, als die komplette Struktur auf einem aktuellen Framework neu zu entwerfen.

Hinsichtlich der verwendeten Skriptsprache PHP wurden mit dem Update von PHP5 auf PHP7 ebenfalls grundlegende Änderungen vollzogen, wodurch einige ältere Bestandteile aus dem Quellcode obsolet wurden und unter PHP7 nicht mehr nutzbar sind. Dies wiederum bricht die Abwärtskompatibilität, sodass älterer Code zeitaufwendig refaktoriert werden muss, um den neuen Spezifikationen zu entsprechen.

Zudem hat die als Betriebssystem verwendete Software Debian 6.0.10 bereits am 29. Februar 2016 das Ende der Langzeitunterstützung erreicht und erhält somit auch keine Sicherheitsupdates mehr.

Die Auswahl geeigneter Softwarekomponenten, die Stabilität und Langzeitsupport gewährleisten, ist somit unerlässlich.

4.2 Datenmodell

Durch die mit der Zeit angepassten Anforderungen der unterschiedlichen Projekte hat sich auch das Datenmodell entsprechend entwickelt. Zu Beginn noch recht überschaubar, expandierte es im Laufe der Jahre und beinhaltete einige Anpassungen, die nicht auf einen künftigen Datenaustausch ausgelegt waren. Neben der Umsetzung grundlegender Datenbank-Konventionen wurden primär die Bedürfnisse der Mitarbeiter berücksichtigt, wodurch die Usability gegenüber einer konsistenten Datenhaltung häufig priorisiert wurde. Dies führte dazu, dass viele Datenfelder des *HSC* als Freitextfelder vorhanden sind und demzufolge komplexe Werte enthalten können, die sinnvollerweise möglichst atomisiert auf mehrere Felder hätten verteilt werden müssen. Daraus ergibt sich die Situation, dass es zwar sehr einfach ist, die Datenbank des HSC mit Daten zu füllen, diese sich aber nur sehr schwer wieder gezielt extrahieren lassen, was aber als Basis für eine effiziente Suche und eine unkomplizierte Datenportierung zwingend erforderlich ist.

Ein Umbau der bestehenden Datenstruktur ist somit unausweichlich und forciert die Frage nach der zukünftigen Datenhaltung. Die derzeit verwendete, auf MySQL basierende relationale Datenverwaltung ist möglicherweise zu unflexibel, um die Tragweite der kommenden Änderungen adäquat umzusetzen, sodass ein Wechsel zu einer dokumentbasierten Datenbank wie MongoDB eine mögliche Alternative bietet. Diese verwendet ein dynamisches Schema, wodurch Änderungen im Datenmodell schneller implementiert werden könnten als in einer relationalen Datenbank, da bisher noch nicht absehbar ist, welche Daten bis 2037 noch in den Bestand des *HSC* aufgenommen und in welcher Form diese vorliegen werden.

4.3 Datennormierung/-strukturänderung im laufenden Betrieb

Der Projektantrag des *HSC* sieht eine Neustrukturierung der Datenbank bzw. des Datenmodells vor. Aufgrund des auf Freitextfeldern basierenden Modells müssen die bereits vorhandenen Daten nachträglich normiert werden, um die Spezifikationen des Projektantrags zu erfüllen. Dieser Mehraufwand eröffnet zwei mögliche Optionen: 1) Parallele Normierung im laufenden Betrieb der aktuellen Datenbank, um die Daten anschließend sauber ins neue Modell zu importieren, oder 2) Mitnahme des alten Datenmodells in das neue Datenmodell, um letzteres möglichst schnell im Produktivbetrieb einsetzen zu können. Die erste Möglichkeit würde es zwar gestatten, das Modell völlig neu aufzusetzen und sauber mit Daten zu befüllen, böte aber auch durch die Weiternutzung der veralteten Software bis zum Systemwechsel ein erhöhtes Betriebsrisiko mit ungewissem Endpunkt. Die zweite Möglichkeit würde zwar den unmittelbaren Systemwechsel forcieren, hätte dann aber auch den Nachteil, dass ein neu veröffentlichtes und produktives Web-Interface intensive Betreuung und Fehlerbehebung nach sich zieht.

4.4 Ressourcenrelation

„It's a difficult problem to distribute expertise among project members."[6]

Letztlich steht und fällt die langlebige Laufzeit eines Projekts mit der Expertise der Beteiligten, was wiederum deren Anzahl zur kritischen Größe macht: Fehlen die Experten im Team oder gestaltet sich der Austausch der Expertise als problematisch, weil alle Beteiligten auf ihrem jeweiligen Gebiet zu spezialisiert sind, kann dies die Projektdauer verlängern oder das Projekt unter gewissen Voraussetzungen (wie z. B. dem Ausfall eines oder mehrerer Teammitglieder) sogar zum Stillstand bringen. Diese Problematik ist im Risikomanagement auch als *Truck Number* bekannt:

„The Truck Number is the size of the smallest set of people in a project such that, if all of them got hit by a truck, the project would be in trouble."[7]

Nach dieser Definition ist ein Projekt mit einer *Truck Number* von 1 höchst gefährdet, da das gesamte Projektwissen in einer einzigen Person gebündelt ist. Fällt diese Person aus, kann das Projekt in der Folge zum Stillstand kommen.

Der *HSC* hat derzeit für den Bereich Digital Humanities eine *Truck Number* von 1. Erschwerend kommt hinzu, dass die derzeitigen Personalstrukturen in den Geisteswissenschaften das Risiko eines Ausfalls u. a. durch knapp geplante Projektspezifika, befristete Verträge und gleichzeitige Qualifikationsvorhaben der wissenschaftlichen Mitarbeiter erhöhen. Somit sind die derzeitigen Voraussetzungen in den geisteswissenschaftlichen Personalstrukturen eher kontraproduktiv für nachhaltige Arbeit in diesem Bereich.

[6] http://wiki.c2.com/?TruckNumber (Abgerufen am 06.09.2018).
[7] http://wiki.c2.com/?TruckNumberFixed (Abgerufen am 06.09.2018).

5. Ausblick/Fazit

Langjährig gewachsene Projekte wie der *HSC*, der in dieser Ausrichtung die inhaltlich diametralen Fachrichtungen der älteren deutschen Literatur und der Informatik in sich vereint, stehen exemplarisch für deren erfolgreiche Kooperation. Obwohl die digitale Erfassung der *HSC*-Daten aus der Sicht der beteiligten Geisteswissenschaftler seinerzeit nicht vorbehaltlos betrachtet wurde, entwickelte sich daraus über die Jahre ein wichtiger Pfeiler für die wissenschaftliche Community und ist seitdem nicht mehr wegzudenken. Die fortwährende und punktuelle Weiterentwicklung des Systems trug ihren Teil dazu bei, die digitale Archivierung historisch bedeutsamer Kulturgüter zur Alltagspraxis der Mitarbeiter zu machen.

Entscheidend für die bisherige positive Entwicklung des Projektes ist jedoch die Schnittstelle zwischen Geisteswissenschaften und Informatik, in Person des ehemaligen Projektmitarbeiters, der beide wissenschaftlichen Pole in sich vereinte – die in Abbildung 2 beschriebene Position „What the customer really needed“. In dieser Funktion wird die Kommunikation zwischen beiden Parteien erleichtert, wodurch Missverständnisse im Vorfeld vermieden werden können und daraus ein ‚gemeinsames Vokabular' für den Arbeitsalltag entsteht. So wies z. B. das im Titel genannte und oft genutzte Bonmot „Das ist nicht ganz trivial...“ als Antwort auf die Frage nach einem neuen Feature die Geisteswissenschaftler regelmäßig darauf hin, dass dessen Umsetzung zwar möglich, aber von der Aufwand-Nutzen-Relation her nicht sinnvoll wäre. Während der ehemalige Digital-Humanities-Projektmitarbeiter allerdings Vermittler und Informatiker in einer Person darstellte, wäre es für den zukünftigen Projektverlauf denkbar, diese Doppelfunktion durch weitere personelle Aufstockung aus dem Informatik-Bereich zu entlasten: Dies würde nicht nur die *Truck Number* erhöhen und somit das Ausfallrisiko minimieren, sondern das Projekt durch die Hinzunahme eines bzw. mehrerer Spezialisten weiter professionalisieren. Mit der Kenntnis beider Disziplinen sowie deren Vokabular würde ein hybrid ausgebildeter *Jack-of-all-trades* somit zum kommunikativen Vermittler zwischen den Welten, während eine höhere Spezialisierung im Informatik-Bereich dem gestiegenen Projektvolumen Rechnung tragen würde.

Es bleibt festzuhalten, dass Projekte wie der *HSC* ohne digitale Strukturen nicht möglich sind, aber in Zukunft vor der großen Aufgabe stehen, die im geisteswissenschaftlichen Kontext bisher aus DFG-Anträgen bekannten kurzlebigen Projektstrukturen durch eine nachhaltige Langzeitplanung an die Anforderungen zeitgemäßer IT-Projekte anzupassen. Dazu ist nicht nur das entsprechende Expertenpersonal erforderlich, sondern auch Mitarbeiter mit hybridem Anforderungsprofil, die zwischen den Disziplinen als kommunikative Schnittstelle fungieren.

Literaturverzeichnis

Göldner, Axel / Rüter, Andreas / Schröder, Jürgen (Hg.): IT-Governance in der Praxis. Berlin/Heidelberg: Springer 2010.

Abbildungsverzeichnis

Abb. 1: Kommunikative Missverständnisse in der Projektarbeit. © http://www.projectcartoon.com/pdf.php?CartoonID=3&PaperSize=A4 (Zugriffsdatum: 06.09.2018).

Abb. 2: Exemplarische Projektvorstellungen von Geisteswissenschaftlern (links) und Informatikern (rechts) gegenüber dem eigentlichen Desiderat.
©http://www.projectcartoon.com/pdf.php?CartoonID=3&PaperSize=A4 (Zugriffsdatum: 06.09.2018).

III. Veränderungen in der digitalen Verbundforschung

Vom Wandel hin zur objektbasierten Forschung im Kontext von DARIAH-DE

Mark Fichtner, Tobias Gradl, Canan Hastik (Nürnberg / Bamberg / Darmstadt)

1. Forschung im Kontext von DARIAH-DE

DARIAH-DE bietet auf Basis der Datenföderationsarchitektur (DFA) eine langfristige und persistente Verfügbarkeit und Nutzbarkeit von Forschungsdaten und -materialien, während gleichzeitig übergreifende Such- und Analysemöglichkeiten in den verzeichneten heterogen strukturierten Forschungsdatensammlungen und Archiven ermöglicht werden. Dadurch nimmt DARIAH-DE bereits jetzt eine besondere Stellung für die digitale Unterstützung von Wissenschaft, Forschung und Lehre in den Geistes- und Kulturwissenschaften ein.

Eine zentrale Herausforderung besteht in der Integration fachspezifischer Forschungsdaten unter Berücksichtigung der speziellen Charakteristika verteilter Forschungsdatensammlungen. Die stark variierenden geistes- und kulturwissenschaftlichen Forschungsdaten und Forschungsfragen erfordern individuelle Gestaltungsmöglichkeiten bei der Entwicklung von Forschungsdatensammlungen. Dies wird insbesondere durch das breite Spektrum an Metadatenformaten deutlich. Die Diversität der Forschungsdaten hat auch Einfluss auf domänenübergreifende Nutzungsmöglichkeiten. Insbesondere die objektbasierte Forschung, wie sie Schwerpunkt an Museen und in der Kunstwissenschaft ist, bringt hier zusätzliche Herausforderungen mit sich: Objektsammlungen können zwar in der Regel mit Datensätzen unterschiedlicher Provenienz durch eine gemeinsame Referenz auf Personendaten, Orte oder andere Typologien und Klassifikationen koreferenziert werden; Objektbeschreibungen zielen jedoch darauf ab, abstrakte Konzepte und komplexe Informationszusammenhänge sowie ereignisbasierte Informationen miteinander zu vernetzen. Der Umgang mit tief erschlossener Information aus objektbasierten Datensammlungen ist essentiell für die nachhaltige Nutzung und Etablierung von Forschungsinfrastrukturen in den Geistes- und Kulturwissenschaften. Letztlich kann nur dadurch die fachübergreifende Vernetzung der unterschiedlichen beteiligten Fachkulturen gefördert und eine transdisziplinäre Zusammenarbeit ermöglicht werden.

Objektdaten können mit Hilfe von WissKI (Wissenschaftliche KommunikationsInfrastruktur) – eines Softwaresystems, das im Rahmen eines DFG-geförderten, gleichnamigen Forschungsprojekts entstanden ist – in tief erschlossener Form abgebildet werden. Das System ermöglicht die Entwicklung von Korpora mit Forschungsprimärdaten und wird darüber hinaus auch für die Vermittlung von Kulturerbe in Form von Ausstellungen eingesetzt. WissKI unterstützt die Speicherung, Verwaltung und Vernetzung fachspezifischer Objektdaten im Sinne von Linked Open Data (LOD). Der Ansatz der DARIAH-DE DFA hingegen ist die Unterstützung eines domänenübergreifenden Forschungsdatenmanagements.

Unter dem Dach der DARIAH-DE DFA sollen im Rahmen einer Machbarkeitsstudie WissKI-basierte Forschungsdaten prototypisch indiziert und zugänglich gemacht werden. Ziel ist es öffentlich zugängliche, fachspezifische Forschungsdatensammlungen zusammenzuführen, um diese langfristig sichtbar, persistent verfügbar, interoperabel zugänglich und nachnutzbar zu machen. Dies umfasst die Registrierung dieser Sammlungen über die DARIAH-DE Collection Registry, die Modellierung und Transformation der Sammlungsdaten mit Hilfe des DARIAH-DE

Data Modeling Environments (DME) sowie die nachhaltige und langfristige Speicherung objektbasierter Forschungsdaten über WissKI direkt im DARIAH-DE Repository. Die Studie adressiert somit die Diskrepanz von objektbasierten, semantischen Daten und deskriptiven, nichtsemantischen Daten sowie damit verbundene wissenschaftliche Sichtweisen und Praktiken. Die entwickelten Anwendungsfälle zur Integration von CIDOC CRM-basierten WissKI-Repositorien mit der DARIAH-DE DFA, die damit verbundene notwendige Reduktion der Information, Auswirkungen auf Suchmöglichkeiten, aber auch Lösungswege und Potentiale werden in diesem Beitrag erläutert und diskutiert.

DARIAH-DE und WissKI treiben somit die Verschränkung von dokumentbasierter und objektbasierter Forschung voran. Der über die DARIAH-DE DFA zugängliche Datenbestand ist im Rahmen der objektbasierten Forschung insbesondere für die Kontextualisierung und damit verbunden für die Sichtbarkeit und Auffindbarkeit der Sammlungsobjekte von besonderer Bedeutung. Die Vernetzung objektbasierter, spezifischer Datenpools fördert somit die sammlungs- und disziplinübergreifende Forschung. Dies kann z.B. durch die Identifikation von relevanten Schnittmengen (z.B. Objekttypus, Orte und Personen im Rahmen der Provenienzforschung) in den bestehenden Datenmengen geschehen – während die DARIAH-DE DFA eine Brückenfunktion einnimmt.

1.1 Forschungsdatenmanagement mit DARIAH-DE

In einer achtjährigen Gesamtförderung entsteht seit 2011 die virtuelle Forschungsinfrastruktur DARIAH-DE als Bestandteil des Gesamtprojektes DARIAH-EU. Eine zentrale Komponente von DARIAH-DE ist die DARIAH-DE Datenföderationsarchitektur (DFA). Eine wesentliche Aufgabe der beinhalteten Komponente besteht darin, strukturierte Daten zu modellieren und zu assoziieren, mit dem Ziel, integrative Sichten auf heterogene Daten zu erstellen und auf diese beispielsweise im Rahmen einer übergreifenden Suche zugreifen zu können[1]. Bislang werden hierfür in der Regel heterogene Datenmodelle und disziplinspezifische Metadatenstandards, wie Text Encoding Initiative (TEI) oder Lightweight Information Describing Objects (LIDO) mit einfachem Dublin Core (DC) assoziiert.

Mit Blick auf die Verarbeitung, Transformation und Integration von Daten aus verteilten, heterogenen Forschungsrepositorien bestehen die wesentlichen Komponenten der DARIAH-DE DFA aus der Collection Registry und dem Data Modeling Environment (DME). Die Generische Suche der DARIAH-DE DFA nutzt die Informationen und Funktionalität dieser sogenannten Föderationsschicht, um Daten zu adressieren und verarbeiten (Abb. 1). Das DARIAH-DE Repository dient als nachhaltiger Speicher für Nutzer, beispielsweise für Forschende ohne Zugang zu einem Repositorium in den eigenen Organisationsstrukturen.

[1] Gradl, Tobias / Henrich, Andreas: Die DARIAH-DE-Föderationsarchitektur – Datenintegration im Spannungsfeld forschungsspezifischer und domänenübergreifender Anforderungen. In: *Bibliothek Forschung und Praxis*, 40, 2 (2016), S. 222-228.

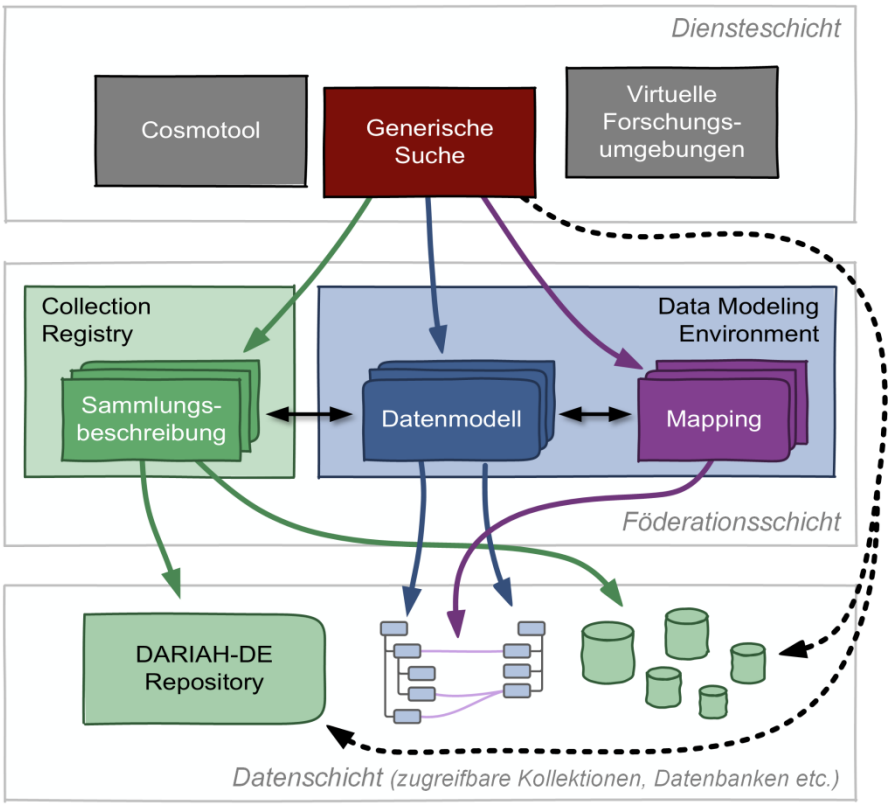

Abb. 1: Generische Suche

1.2 Virtuelle Forschungsinfrastruktur WissKI

Das DFG-finanzierte Projekt "WissKI"[2] hat in den Jahren 2009 bis 2011 zur Entwicklung einer virtuellen Forschungsumgebung für die Anwendung im Bereich der Digital Humanities geführt. In einer zweiten Förderungsphase wurde die Forschungsumgebung in den Jahren 2014 bis 2016 von den Projektpartnern, der Friedrich-Alexander-Universität Erlangen-Nürnberg, dem Germanischen Nationalmuseum sowie dem Zoologischen Forschungsmuseum Alexander Koenig konsequent erweitert. Hauptaspekt der Datenerfassung und -haltung in WissKI sind die semantischen Zusammenhänge durch umfassende Unterstützung aktueller Semantic Web-Technologien. Die Einordnung und Speicherung der erhobenen Daten erfolgt auf Grundlage des ISO-Standards 21127 (CIDOC CRM)[3]. Auf Basis dieser Technologie werden solitär erscheinende Daten zu einem gemeinsamen, semantischen Netzwerk verbunden, um damit die unmittelbare Sichtbarkeit weiterer, tiefergehender Zusammenhänge zu ermöglichen. Daraus ergibt sich ein Mehrwert, der in der Vergangenheit in flachen Hierarchien wie Datenbanktabellen gar nicht oder nur mit sehr hohem Aufwand erkennbar gemacht werden konnte. Das webbasierte Systemdesign und der dadurch ermöglichte Zugriff über das Internet, die Anbindung von extern kuratierten Datenquellen (sog. Authority Files) und die Möglichkeit zur Bereitstellung ausgewählter Daten über gängige Online-Schnittstellen (Web-Frontend, SPARQL-Endpoint, OAI-PMH, etc.) betonen den Semantic Web-Gedanken hinter dieser Infrastruktur.

[2] vgl. http://risources.dfg.de/detail/RI_00397_de_druck.pdf (Zugriffsdatum: 16.12.2018).
[3] vgl. http://cidoc-crm.org (Zugriffsdatum: 16.12.2018).

2. Verschränkung von dokument- und objektbasierter Forschung

In den meisten Disziplinen der Geisteswissenschaften spielen Dokumente und Textformen eine elementare Rolle. Ausgehend von den Primärquellen hat sich eine an Volltexten und verschiedenen Textsorten orientierte Forschungspraxis etabliert. Auch die Forschungsergebnisse hatten in der Vergangenheit üblicherweise Textform und erschienen als Aufsatz, Manuskript oder Buch. Deskriptive Metadaten werden in diesem Zusammenhang eingesetzt und genutzt, um den Text bestmöglich zu beschreiben und um eine maschinengestützte Suche zu gewährleisten. Die standardisierte und gleichzeitig qualitativ hochwertige Erschließung der Texte ermöglicht Analyse- und Aggregationsverfahren bezogen auf einzelne Dokumente oder das Gesamtkorpus. Diese Anfragen können zusätzlich eingeschränkt, sortiert und gruppiert werden. Genau diese Forschungspraxis spiegelt sich in den meisten Forschungsdaten der DARIAH-DE DFA zum derzeitigen Stand wider.

Davon unterscheidet sich die objektbasierte Forschung. Bei ihr stehen die Gegenstände selbst im Fokus der Forschungsarbeit. Sie werden anhand ihrer materiellen Beschaffenheit, hinsichtlich ihrer Entstehung, Funktion und Wirkungsgeschichte erschlossen und geographisch, temporal und kulturgeschichtlich eingeordnet. Oft sind es dabei die impliziten und nicht unmittelbar sichtbaren Eigenschaften und Bezüge eines Sammlungsobjektes mit anderen Entitäten des Diskurses, die seine Bedeutung oder die eigentliche Forschungsleistung ausmachen. Für die Modellierung von Daten im kunst- und kulturwissenschaftlichen Bereich hat sich das CIDOC Conceptual Reference Model (CRM), ISO-Standard 21127, herausgegeben von der Arbeitsgruppe für Dokumentation der ICOM, durchgesetzt. Das CIDOC CRM ist eine formale Ontologie, die die Integration, Vermittlung und den Austausch heterogener Informationen des kulturellen Erbes in der Dokumentation ermöglichen soll. Für den Bereich des Kulturerbes entwickelt, soll das CIDOC CRM als gemeinsame “Lingua Franca” zwischen den verschiedenen Wissenschaftsbereichen dienen.

Die unterschiedliche Forschungspraxis bedingt nicht zuletzt eine unterschiedliche Datenhaltung, die nur schwierig in einem gemeinsamen Portfolio vereint werden kann. Dokumentenbasierte und objektbasierter Datenhaltung unterscheiden sich im Wesentlichen hinsichtlich der an sie gestellten Anforderungen. Während dokumentenbasierte Datenhaltung versucht, sämtliche Information im Dokument zu kapseln und externe Information zu referenzieren, spielt bei der objektbasierten Datenhaltung der Kontext eine so gewichtige Rolle, dass er zumeist separat miterfasst wird. So sind beispielsweise Personendaten, Ortsinformation, Ereignisdaten, Zeitangaben und ähnliche Objektbezüge typisches Beiwerk. Eine zeitgemäße Forschung fordert jedoch einen einheitlichen Zugriff auf Information, eine nach Möglichkeit einheitliche, persistente Speicherung sowie tiefgreifende Erschließung mit langfristiger Interpretierbarkeit der Information, sowohl von dokumentbasierter wie auch objektbasierter Datenhaltung. An moderne Forschungsdatenföderationsinfrastrukturen wie die DARIAH-DE DFA werden die Anforderungen gestellt, diese Bedarfe zu erfüllen.

2.1 Vernetzung heterogener Datensammlungen in DARIAH-DE

Die Einbindung von Datenbeständen in die Infrastruktur von DARIAH-DE erfolgt in einem dreischrittigen Verfahren: Diese Schritte werden exemplarisch am Medienarchiv zur Fränkischen Tafelmalerei von Dürer[4] umgesetzt, um die Integration von Daten aus einer auf WissKI basierenden virtuellen Forschungsinfrastruktur zu veranschaulichen. Zunächst wird (1) hierfür in der Collection Registry das Medienarchiv als Sammlung verzeichnet und ein Verweis auf die verfügbare OAI-PMH Schnittstelle eingetragen[5]. Die auf diese Weise aus dem Medienarchiv exportierten Daten werden (2) anschließend in der DME modelliert[6]. Hier kann nun eine Anreicherung bzw. Transformation der Daten stattfinden, so können z.B. kontextspezifische, nicht in den Daten enthaltene Informationen, wie weiterführende Informationen zu den Personen oder direkte Links zu den Bilddaten ergänzt werden. Schließlich werden (3) verschiedene Mappings zu den Datenmodellen[7,8] angelegt, die für den Betrieb der Generischen Suche Voraussetzung sind. Das zur Verfügung gestellte Medienarchiv der Fränkischen Tafelmalerei von Dürer kann dadurch mit Inhalten anderer über die DARIAH DFA vernetzten Sammlungen ergänzt werden (Abb. 2).

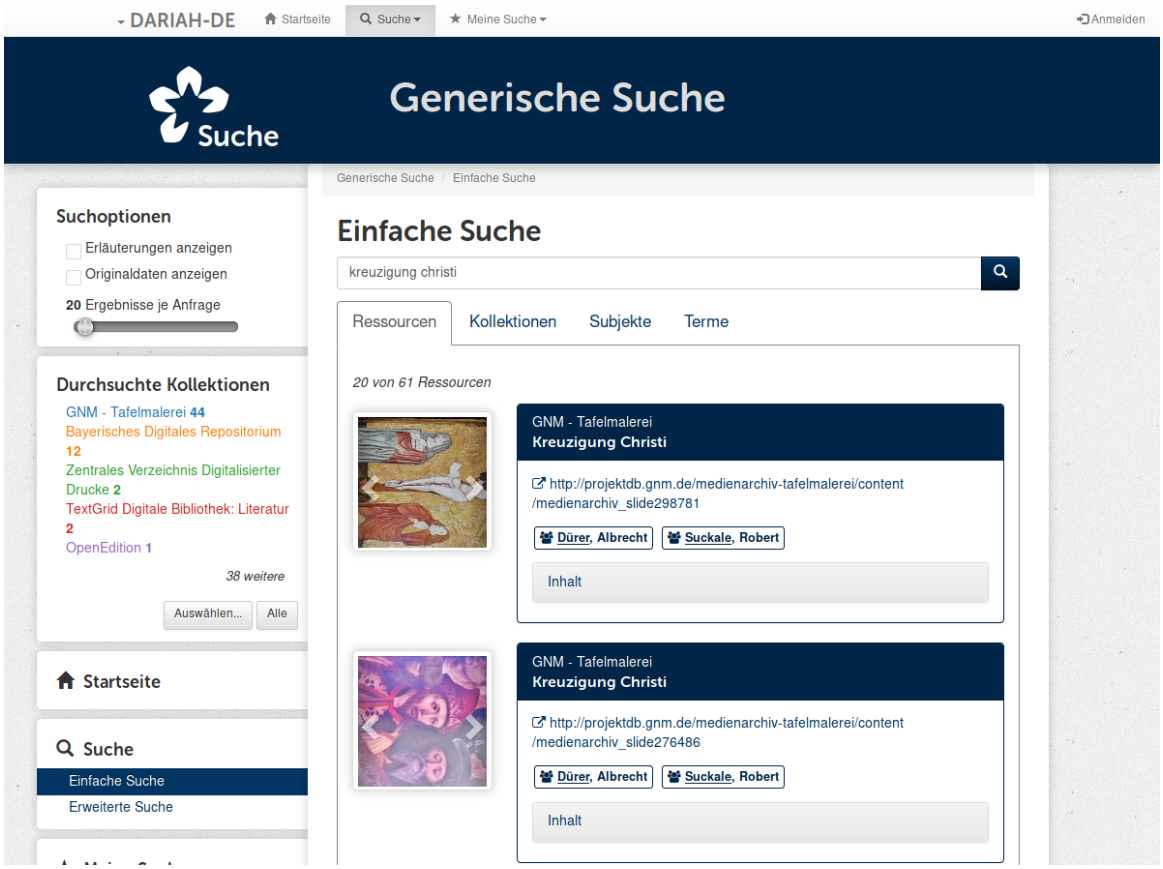

Abb. 2: Ergebnisse aus dem Medienarchiv Tafelmalerei des GNM über die Generische Suche

[4] http://projektdb.gnm.de/medienarchiv-tafelmalerei (Zugriffsdatum: 16.12.2018).

[5] https://colreg.de.dariah.eu/colreg-ui/collections/58a57d232d46f37c25682a87 (Zugriffsdatum: 16.12.2018).

[6] https://dme.de.dariah.eu/dme/model/editor/58a5b5fae4b0d67f5ba61da1 (Zugriffsdatum: 16.12.2018).

[7] https://dme.de.dariah.eu/dme/mapping/editor/5b86742f61432f13a8fd73e7 (Zugriffsdatum: 16.12.2018).

[8] https://dme.de.dariah.eu/dme/mapping/editor/5b8fea3561432f44d6419766 (Zugriffsdatum: 16.12.2018).

Das von WissKI genutzte CIDOC CRM ist ein Novum im Rahmen der DARIAH DME, weshalb ein entsprechendes Mapping zu erarbeiten ist. Nachdem dieses erarbeitet wurde, entstehen auf der Basis der neu dazugewonnen vernetzten Sammlungen im Wesentlichen drei Anwendungsfälle:

Sichtbarkeit, Nachnutzbarkeit und Kontextualisierung: Eine Vielzahl neuer Sammlungen, insbesondere aus dem musealen Kontext, können über die DARIAH-DE Infrastruktur langfristig und persistent auffindbar und zugänglich gemacht werden.

Forschungs- und projektspezifische Integration: Heterogene (Meta-)Daten aus unterschiedlichen Kontexten können im Hinblick auf spezifische Informationsbedürfnisse des jeweiligen Nutzerkreises aufbereitet und möglichst verlustfrei zusammengeführt werden. Hierfür können im Rahmen der Nutzerszenarien auf bestehende Modelle und Mappings zurückgegriffen und diese erweitert werden.

Migration und Zusammenführung von Daten: Bestehende DARIAH Sammlungsbestände können durch WissKI in Semantic Web konforme Darstellungen transformiert und dort verwaltet werden. Die Werkzeuge der DARIAH-DE DFA können das hierfür notwendige Mapping auf CIDOC CRM und die Kontextualisierung sowie Anreicherung unterstützen. Hierfür ist ein dedizierter Transformationsdienst notwendig, dessen Implementierung ab Ende 2019 zur Verfügung stehen soll. Dieser soll den DARIAH Nutzerkreis insbesondere bei der (ggf. wiederholt vorzunehmenden) Migration von Daten unterstützen.

Aufgrund der Breite des DARIAH-DE Nutzerspektrums und der dadurch bedingten Anwendungsvielfalt wurde bei der Entwicklung der DFA auf die Definition statischer Integrationsmodelle von vornherein verzichtet. Stattdessen werden in der Infrastruktur zugängliche Daten und Metadaten in möglichst originärer Form dokumentorientiert mit Hilfe hinterlegter Mappings modelliert zusammengeführt. Dies hat insbesondere den Vorteil, dass Daten in Abhängigkeit konkreter Anwendungskontexte interpretiert und so auch konfliktäre, sich überschneidende Perspektiven auf Daten abgebildet werden können. Ein Beispiel hierfür ist die Nachnutzung von Referenzen auf Personendaten, die je nach Anwendungskontext unterschiedliche Informationsbedürfnisse bedienen sollen, welche wiederum Einfluß auf die Verarbeitung und Interpretation dieser Daten haben. Zwar wird, wie oben beschrieben, in der Hauptinstanz der Generischen Suche derzeit Dublin Core als de-facto Standard zur Integration verwendet. Spezifische Installationen der Generischen Suche, wie z. B. die des Forschungsverbundes Marbach Weimar Wolfenbüttel, nutzen jedoch komplexere oder auch mehrere Datenmodelle.

2.2 Showcase: Albrecht Dürer

Im Rahmen des ersten Anwendungsfalls der Machbarkeitsstudie werden zunächst die Aspekte Sichtbarkeit, Nachnutzbarkeit und Kontextualisierung von Daten adressiert und die ersten Ergebnisse präsentiert**.** Hierfür wird ein Mapping erstellt (Abb. 3), in dem u.a. Creator, Publisher und Contributor-Elemente den CIDOC CRM Entitäten E21 (Person) und E74 (Group) zugeordnet werden. Die tatsächliche Gestaltung der Entitäten und deren Eigenschaften erfolgen auf Instanzebene nach spezifischen Regeln, die sich je nach Sammlung unterscheiden und beliebig komplex sein können.

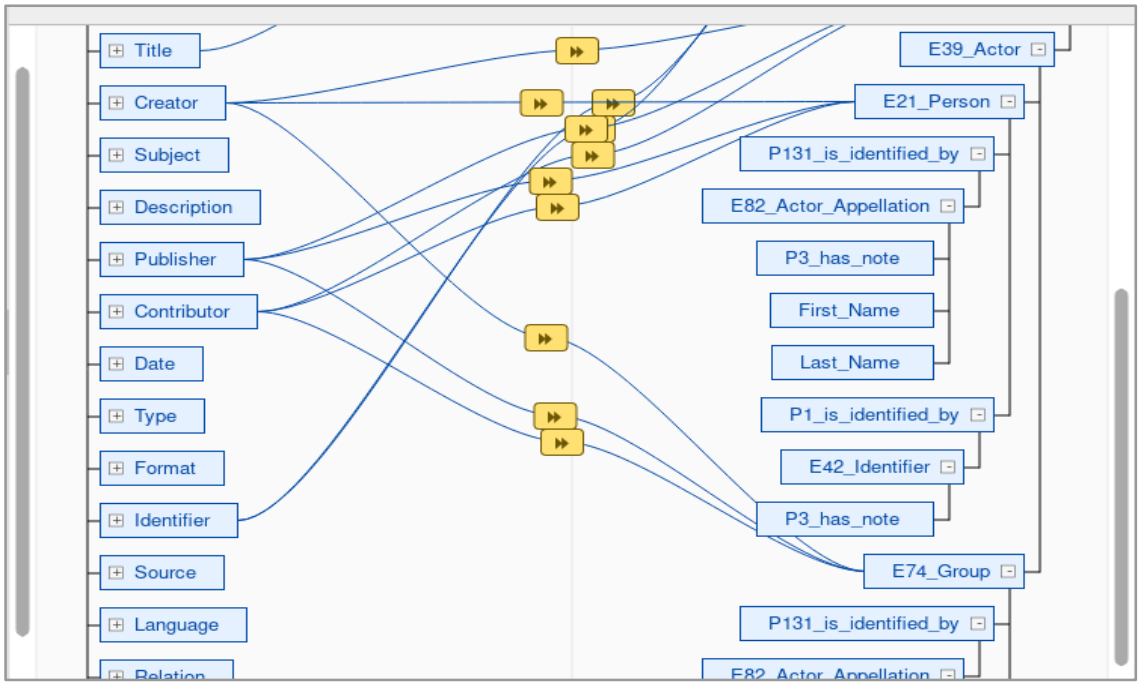

Abb. 3: Mapping auf CIDOC CRM

Übertragen auf die Generische Suche kann die Auszeichnung von Personen zur Aggregation von Dokumenten und zur Darstellung herangezogen werden. Neben der Suche nach einzelnen Personen ermöglicht die Generische Suche derzeit die Aggregation von Suchergebnissen auf Basis von Termhäufigkeit (Abb. 4). Suchende können anhand dieser einfachen Zusammenfassungen eine erste Beurteilung der Relevanz von Sammlungen für die eigene Forschung vornehmen und ihre vertiefenden Recherchen ggf. dort fortsetzen.

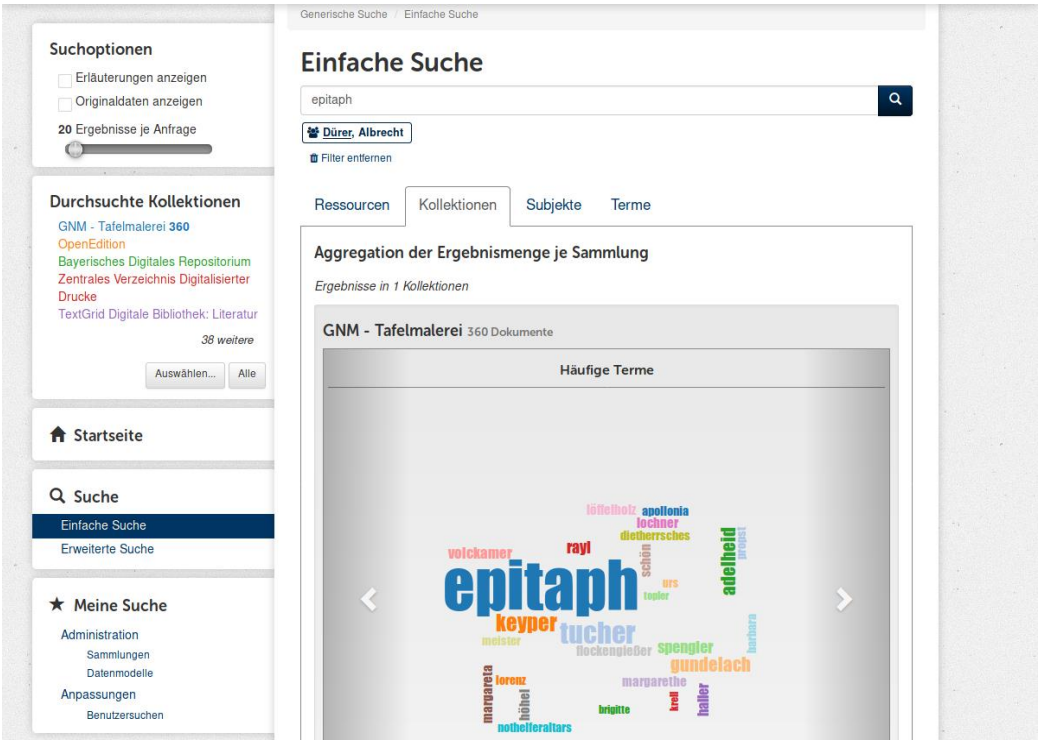

Abb. 4: Häufig auftretende Begriffe in dem Medienarchiv Tafelmalerei des GNM

Weitere Formen der Aggregation sind geplant oder befinden sich derzeit bereits in der Umsetzung. Mit einer fortschreitenden Fokussierung auf objektbasierte Zusammenhänge in den durchsuchten Beständen wird die Umsetzung weiterer Analyse- und Visualisierungstechniken, wie beispielsweise die geotemporale Zusammenfassung von Suchergebnissen je Sammlungsbestand erleichtert oder überhaupt ermöglicht.

Abb. 5: Termhäufigkeit im Medienarchiv des GNM zu Albrecht Dürer

Abb. 6: Termhäufigkeit in Wolfenbütteler Digitalen Bibliothek zu Albrecht Dürer

Die Suche nach der Person Albrecht Dürer in der Generischen Suche ermittelt so beispielsweise aus den knapp 10k Treffern im Medienarchiv zur Fränkischen Tafelmalerei eine Menge an relevanten Termen und verdeutlicht den Kontext der Treffermenge (Abb. 5).
Im Vergleich hierzu weisen die Terme zur Treffermenge in der Wolfenbütteler Digitalen Bibliothek, ebenfalls ein an die DFA angebundenes Archiv, eine deutlich geringere inhaltliche Kohärenz auf, trotz oder gerade wegen der im Vergleich zur Tafelmalerei geringeren Anzahl von Suchergebnissen (Abb. 6).

Die Modellierung und Vernetzung von Personen ermöglicht neben deren Einsatz zur Formulierung und Verfeinerung von Suchanfragen ebenso eine Zusammenfassung von Suchergebnissen anhand der Subjekt-Kookkurrenzen. Diese Form der Aggregation wurde erst kürzlich umgesetzt und wird in Zukunft eine geeignetere Form der Visualisierung finden. Damit ist es möglich, auf Basis des aktuellen Standes, Rückschlüsse über die Suchergebnisse in dem jeweiligen Datenbestand zu treffen.
Die beiden Darstellungen (Abb. 7 und Abb. 8) vergleichen die Treffermenge zur unveränderten Suchanfrage nach der Person Albrecht Dürer in der Wolfenbütteler Digitalen Bibliothek (Abb. 7) und dem Bayerischen Digitalen Repositorium (Abb. 8). Diese Termwolken veranschaulichen eine unterschiedliche Zusammensetzung und in beiden Fällen eine Zerstreuung des Personenfeldes.

Abb. 7: Assoziierte Personen zu Albrecht Dürer in der Wolfenbütteler Digitalen Bibliothek

Abb. 8: Assoziierte Personen zu Albrecht Dürer in dem Bayerischen Digitalen Repositorium

3. Ergebnisse und weitere Schritte

Die Kooperation von DARIAH-DE und WissKI verfolgt die Zielsetzung, die semantische Erschließung und Dokumentation von Kulturobjekten sowie die Anbindung domänenspezifischer Forschungssammlungen weiter auszubauen und zu befördern. Hierdurch wird dem steigenden Bedarf an integrierten heterogenen Datensammlungen, der durch zunehmende vernetzte und disziplinübergreifende geistes-, kultur- und kunstwissenschaftlicher Forschung entsteht, entsprochen und nachgekommen.

Während WissKI nicht nur Datenlieferant ist, sondern auch als Impulsgeber für die Verschmelzung dokumentbasierter und objektbasierter Datenrepositorien wirkt, übernimmt die DARIAH-DE DFA die gewünschte Brückenfunktion zwischen den in Bezug auf die Datenqualität hochwertigen dokument- und objektorientierten Sammlungen sowie wenig strukturierten, für Forschende ggf. nicht weniger relevanten Forschungsdaten. Unsicherheiten mit Blick auf eine mögliche objektbasierte Vernetzung einzelner Ressourcen sind aufgrund der stark variierenden Datenqualität von besonderer Bedeutung für die DARIAH-DE DFA. Eine Modellierung und Auswertung solcher Unsicherheiten abhängig von konkreten Forschungsinteressen individualisierbar sein. Insbesondere Anfragen nach einer möglichst großen Zahl potentieller Zusammenhänge ermöglichen das Auffinden ggf. bislang unbekannter Zusammenhänge, erfordern jedoch eine kritische Auseinandersetzung mit den Suchergebnissen. Die Reduzierung eines ausgegebenen Netzwerks auf Basis valider Assoziationen führt dagegen zu einer präziseren Darstellung, die für eine Anwendung im Rahmen von Breitensuchen ggf. besser geeignet ist.

Erste funktionale Fortschritte der DFA und der Generischen Suche von DARIAH-DE durch die dedizierte Berücksichtigung der Objektebene verdeutlichen den Mehrwert auch für eine Föderation stark heterogener Daten. In der Generischen Suche stellt die Ausweitung des Objektmodells auf Basis von CIDOC-CRM damit ebenso nächste Schritte dar, wie die direkte Nachnutzung neuer Entitäts- und Verbindungstypen in Form geeigneter Such-, Navigations- und Visualisierungstechniken. Da die Umsetzung entsprechender Funktionalität in der Föderationsschicht der DFA (vgl. Abb. 1) erfolgt, steht diese neben der Generischen Suche auch für weitere Szenarien zur Verfügung und kann so perspektivisch auch die Migration von Daten in semantisch reichhaltigere Formen unterstützen.

Literaturverzeichnis

CIDOC CRM Conceptual Reference Model. http://cidoc-crm.org (Zugriffsdatum: 16.12.2018).

DARIAH-DE Digitale Forschungsinfrastruktur für die Geistes- und Kulturwissenschaften. https://de.dariah.eu (Zugriffsdatum: 16.12.2018).

DARIAH-DE DME Datenmodell-Editor. https://dme.de.dariah.eu/dme/model/editor/58a5b5fae4b0d67f5ba61da1 (Zugriffsdatum: 16.12.2018).

DARIAH-DE Mapping-Editor. https://dme.de.dariah.eu/dme/mapping/editor/5b86742f61432f13a8fd73e7 (Zugriffsdatum: 16.12.2018).

DARIAH-DE Mapping-Editor. Tafelmalerei. https://dme.de.dariah.eu/dme/mapping/editor/5b8fea3561432f44d6419766 (Zugriffsdatum: 16.12.2018).

DARIAH-DE Sammlungseditor. https://colreg.de.dariah.eu/colreg-ui/collections/58a57d232d46f37c25682a87 (Zugriffsdatum: 16.12.2018).

DARIAH-EU. https://www.dariah.eu (Zugriffsdatum: 16.12.2018).

DCMI: Dublin Core Metadata Initiative. Dublin Core Metadata Element Set, Version 1.1: Reference Description. http://dublincore.org/documents/dces (Zugriffsdatum: 16.12.2018).

LIDO - Lightweight Information Describing Objects. XML Schema for Contributing Content to Cultural Heritage Repositories. Version 1.0. http://www.lido-schema.org/schema/v1.0/lido-v1.0-schema-listing.html (Zugriffsdatum: 16.12.2018).

RIsources: Wissenschaftliche KommunikationsInfrastruktur (WissKI). http://risources.dfg.de/detail/RI_00397_de_druck.pdf (Zugriffsdatum: 16.12.2018).

TEI: Text Encoding Initiative. http://www.tei-c.org (Zugriffsdatum: 16.12.2018).

Gradl, Tobias / Henrich, Andreas: Die DARIAH-DE-Föderationsarchitektur – Datenintegration im Spannungsfeld forschungsspezifischer und domänenübergreifender Anforderungen. In: *Bibliothek Forschung und Praxis*, 40, 2 (2016), S. 222-228.

WissKI. http://wiss-ki.eu (Zugriffsdatum: 16.12.2018).

WissKI Medienarchiv Tafelmalerei. http://projektdb.gnm.de/medienarchiv-tafelmalerei (Zugriffsdatum: 16.12.2018).

Abbildungsverzeichnis

Abb. 1: Generische Suche © DARIAH-DE.

Abb. 2: Ergebnisse aus dem Medienarchiv Tafelmalerei des GNM über die Generische Suche © eigene Grafik, 2018.

Abb. 3: Mapping auf CIDOC CRM © eigene Grafik, 2018.

Abb. 4: Häufig auftretende Begriffe in dem Medienarchiv Tafelmalerei des GNM © eigene Grafik, 2018.

Abb. 5: Termhäufigkeit im Medienarchiv des GNM zu Albrecht Dürer © eigene Grafik, 2018.

Abb. 6: Termhäufigkeit in Wolfenbütteler Digitalen Bibliothek zu Albrecht Dürer © eigene Grafik, 2018.

Abb. 7: Assoziierte Personen zu Albrecht Dürer in der Wolfenbütteler Digitalen Bibliothek © eigene Grafik, 2018.

Abb. 8: Assoziierte Personen zu Albrecht Dürer in dem Bayerischen Digitalen Repositorium © eigene Grafik, 2018.

Virtuell Zusammenwachsen: Konzeption, Aufbau und Intention der digitalen Forschungsinfrastruktur im Forschungsverbund MWW

Swantje Dogunke, Timo Steyer (Leipzig / Wolfenbüttel)

1. Einleitung

Der 2014 gegründete und vom BMBF geförderte Forschungsverbund Marbach Weimar Wolfenbüttel (MWW) vereint mit dem Deutschen Literaturarchiv Marbach (DLA), der Klassik Stiftung Weimar (KSW) und der Herzog August Bibliothek Wolfenbüttel (HAB) drei bedeutende Zentren geisteswissenschaftlicher Forschung in Deutschland.[1] Durch die Aktivitäten des Verbunds sollen die einzigartigen Sammlungen der Einrichtungen noch stärker in den Fokus der nationalen und internationalen Wissenschaft gerückt werden und damit auch zu einer erhöhten Sichtbarkeit der Forschungsaktivitäten beitragen. Die Verbundstruktur bietet für die genannten Einrichtungen die Möglichkeit, durch die Bündelung von Ressourcen und Kompetenzen eine digitale Infrastruktur auf Basis gemeinsamer Entwicklung zu schaffen, welche der Sammlungsforschung neue Möglichkeiten und Impulse bietet und die einzigartigen (digitalen) Sammlungen der Verbundeinrichtungen mit Tools und Services aus dem Bereich der *Digital Humanities* in einem virtuellen Forschungsraum zusammenbringt.

Durch umfangreiche Digitalisierungsinitiativen in den beteiligten Häusern wurde erfolgreich die digitale Transformation der analogen Sammlungen begonnen und bis heute fortgesetzt. Diese Transformation umfasst mittlerweile weit mehr als die reine Imagedigitalisierung - Ziel ist es, auch um neue Methoden der *Digital Humanities* zu unterstützen, die Sammlungen in Form von maschinenlesbaren und interoperablen Daten in strukturierter Form bereitzustellen.[2] Der dadurch für die Forschung generierte Mehrwert wird bei MWW paradigmatisch durch bestandsorientierte Forschungsprojekte demonstriert.[3] Diese Vorhaben werden begleitet und umgesetzt durch die digitale Infrastruktur des Verbundes, welche die Potentiale der *Digital Humanities* für die Sammlungsforschung nachhaltig nutzbar macht.[4]
Der digitalen Infrastruktur kommt eine Schlüsselposition zu, in ihr werden nicht nur die Sammlungen und Forschungsaktivitäten von MWW vereint, sondern für die sich aus dem digitalen

[1] vgl. http://mww-forschung.de (Zugriffsdatum: 3.1.2019).
[2] Zur Relevanz maschinenlesbarer Daten vgl. Stäcker, Thomas: Noch einmal: Was sind geisteswissenschaftliche Forschungsdaten? In: DHd-Blog. 6. Dezember 2015. https://dhd-blog.org/?p=5995 (Zugriffsdatum: 3.1.2019).
[3] Insgesamt umfasst MWW neun bestandsbezogene Forschungsprojekte. Vgl. https://www.mww-forschung.de/forschungsprojekte (Zugriffsdatum: 3.1.2019). Die Erschließung und Auswertung eines Bestands unter Verwendung von DH-Methoden ist für das Projekt "Frühneuzeitliche Gelehrtenbibliotheken" in einem Aufsatz thematisiert worden. Vgl. Beyer, Hartmut / Münkner, Jörn / Schmidt, Katrin / Steyer, Timo: Bibliotheken im Buch: Die Erschließung von privaten Büchersammlungen der Frühneuzeit über Auktionskataloge. In: Hannah Busch / Franz Fischer / Patrick Sahle [Hg.]: Kodikologie und Paläographie im digitalen Zeitalter 4 (Schriften des Instituts für Dokumentologie und Editorik, 11). Norderstedt: BoD 2017, S. 43-70.
[4] Der Begriff Infrastruktur ist in diesem Kontext missverständlich, da er sowohl Hardware und Systeme aber auch Webservices und Informationsangebote umfasst. Infrastruktur ist in diesem Beitrag auf alle im Rahmen von MWW aufgebauten Komponenten der Forschungsumgebung (Speicher, Forschungsraum, Metadatengrammatik) bezogen und dezidiert auf den Aspekt des dauerhaften Betriebes ausgelegt. Generell ist das Vokabular im Bereich der digitalen Geisteswissenschaften häufig nicht ausreichend definiert und zeichnet sich durch Ambiguität aus. So bezeichnen die Begriffe Forschungsportal, Forschungsraum und Forschungsumgebung nicht klar voneinander trennbare Angebote.

Wandel resultierenden Anforderungen werden kooperative und integrative Lösungen angestrebt, welche jede der beteiligten Einrichtungen nicht oder nur eingeschränkt alleine umsetzen könnte.[5] Die MWW-Infrastruktur setzt dabei auf den vorhandenen Daten der beteiligten Häuser auf und bietet neue Services und Tools für das wissenschaftliche Arbeiten mit diesen Daten an. In diesem Kontext ist die Infrastruktur als eine Art *Middleware* zu betrachten, die zwischen der bestehenden Infrastruktur der Häuser und den generierten Daten auf der einen Seite und den wissenschaftlich arbeitenden Nutzern/Innen auf der anderen Seite lokalisiert ist.

Zu Beginn des Forschungsverbundes war die Konzeption der digitalen Infrastruktur nur in ihrer strategischen Ausrichtung festgelegt. Folgende Komponenten wurden identifiziert und im Antrag umrissen:

- Ein virtueller Forschungsraum für die bestandsbezogene Forschung, der entsprechend der Definition der Allianz der deutschen Wissenschaftsorganisationen „eine kooperative Forschungstätigkeit durch mehrere Wissenschaftlerinnen und Wissenschaftler an unterschiedlichen Orten zu gleicher Zeit ohne Einschränkungen ermöglicht.“[6]
- Ein verlässlicher Speicher, um eine Langzeitverfügbarkeit der Daten zu gewährleisten und so den Forscher/Innen eine verlässliche Basis für die wissenschaftliche Nutzung der Sammlungen im Forschungsraum ebenso garantiert wie die nachhaltige Verfügbarkeit der ihrer generierten Forschungsdaten.
- Eine Metadatengrammatik, in der die vorhandenen Sammlungen und erhobenen Forschungsdaten für die Verwendung im Forschungsraum aufbereitet und angereichert werden.[7] Dafür wurde sich verbundweit auf Standards und Anwendungsprofile für die unterstützten (Meta-)Datenformate und verwendeten Normdaten verständigt sowie ein speziell auf die Bedürfnisse der MWW-Infrastruktur zugeschnittenes Metadatenformat entwickelt. Ebenso bilden technische und administrative Metadaten das Rückgrat für das Zusammenspiel der einzelnen Funktionen des virtuellen Forschungsraums und der Vernetzung mit den vorhandenen Infrastrukturen.

[5] Wie groß die zukünftigen Anforderungen durch den digitalen Wandel vor allem in den Bereichen der Automatisierung und künstlichen Intelligenz für Bibliotheken sein könnten, vermittelt der NMC Horizon Report 2017 in der Library Edition. Vgl.: The New Media Consortium (Hg.): NMC Horizon Report. 2017 Library Edition. http://cdn.nmc.org/media/2017-nmc-horizon-report-library-EN.pdf (Zugriffsdatum: 3.1.2019).

[6] vgl. Arbeitsgruppe Virtuelle Forschungsumgebungen in der Allianz der deutschen Wissenschaftsorganisationen (Hg.): Definition Virtuelle Forschungsumgebung, 2011. DOI: http://doi.org/10.2312/ALLIANZOA.028 (Zugriffsdatum: 3.1.2019).

[7] Zur Relevanz von Metadaten für die Forschung vgl. Kramer, Michael J.: Going Meta on Metadata. In: Journal of Digital Humanities, 3, 2 (2014). http://journalofdigitalhumanities.org/3-2/going-meta-on-metadata/ (Zugriffsdatum 6.8.2018).

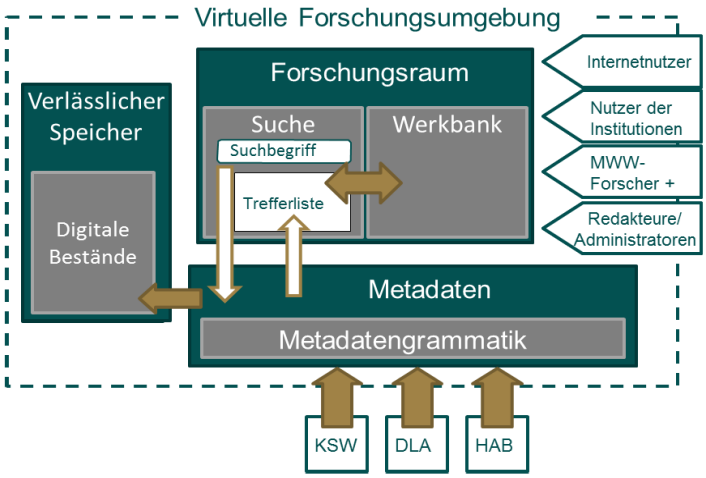

Abb. 1: Erste Architekturskizze der Virtuellen Forschungsumgebung MWW

Der vorliegende Beitrag konzentriert sich auf den virtuellen Forschungsraum und die Metadatengrammatik und thematisiert, welche Faktoren die Konzeption und Umsetzung des Virtuellen Forschungsraum MWW beeinflusst haben und welche Erfahrungen im Kontext von forschungsorientierten, digitalen Infrastrukturaufbau gemacht wurden. Die Perspektiven, die dabei eingenommen werden, sind erstens die Ansprüche der Nutzer/Innen, zweitens die Potentiale und Limitationen durch die vorhandenen Daten und Workflows. Der Beitrag versteht sich als Werkstattbericht, in dem die Erfahrung aus fünf Jahren Entwicklungsarbeit wiedergeben wird. Es wird daher weniger um technische Details gehen, sondern vielmehr der Frage nachgegangen, wie Erwartungen und Ansprüche an digitale Infrastruktur mit der konkreten Entwicklung und dem Ergebnis korrespondieren. Dabei wird auch den erreichten Grad an Akzeptanz bzw. die erforderliche Legitimation eingegangen, welche für die Integration von *Digital Humanities* Methoden in bestandsbezogenen Forschungsprojekte erreicht werden musste.

Wichtig für das weitere Verständnis ist die Struktur der MWW-Projekte. Insgesamt gibt es drei DH-Projekte im Verbund mit je einem wissenschaftlichen Mitarbeiter/In, die jeweils für eine der Komponenten - virtueller Forschungsraum, Metadatengrammatik und Verlässlicher Speicher - verantwortlich sind.[8] Die DH-Mitarbeiter/Innen unterstützten aber auch die bestandsbezogenen Forschungsprojekten in Fragen zu Methoden und Tools aus dem Bereich der *Digital Humanities*. Durch diese Kombination wirkten die Forschungsprojekten auch impulsgebend auf die Entwicklung des virtuellen Forschungsraums.

Der erste Schritt in der Konzeption der digitalen Infrastruktur konzentrierte sich auf die Ermittlung der Bedarfe der Nutzer/Innen, einer Analyse der vorhandenen Datenbestände sowie einer Untersuchung der bestehenden Infrastruktur der Häuser, um eine erste Skizze der häuserübergreifenden Verbundinfrastruktur zu entwerfen (s. **Fehler! Verweisquelle konnte nicht gefunden werden.**). Es wurde deutlich, dass der gemeinsame Aufbau einer digitalen Forschungsinfrastruktur nur durch eine Bündelung der Kompetenzen und enger Zusammenarbeit zwischen den DH-Mitarbeiter/Innen an den drei Standorten zu realisieren ist. Zusätzlich muss-

[8] vgl. https://www.mww-forschung.de/digitale-forschungsinfrastruktur/ (Zugriffsdatum: 3.1.2019).

ten gemeinsame Organisationsformen und Strukturen für die Durchführung und das Management der Projekte aufgebaut werden. Darauf aufbauend wurde ein gemeinsamer Arbeitsplan erstellt, um den Projektfortschritt zu planen und zu dokumentieren.[9]

2. Analyse der Anforderungen der Nutzer/Innen

Die initiale Bedarfsanalyse stützte sich auf den exzerpierten Anforderungen des genehmigten Antrages, qualitativen Interviews der Antragsteller/Innen sowie den Vorgaben des Fördergebers. Zunächst galt es, diese Informationen zu nutzen, um den Nutzerkreis zu erkennen und genauer zu beschreiben.

Es wurden drei Nutzerkreise identifiziert:

1. die MWW-Mitarbeiter/Innen aus den bestandsbezogenen Forschungsprojekten,
2. Forscher/Innen aus den beteiligten Institutionen und Verantwortliche für die bestehende Forschungsinfrastruktur sowie
3. eine Gruppe potentieller Nutzer/Innen, die bereits im Bereich der *Digital Humanities* forschen und nicht an den Institutionen angestellt sind.

Es folgten qualitative und quantitative Befragungen, um Bedarfe zu ermitteln, zu bewerten und zu konkretisieren. Das Ergebnis der Befragung zeigte ein sehr heterogenes Bild, wobei vier Ausprägungen identifiziert werden konnten, die in jedem der genannten Nutzerkreise auftauchten und deren Einbindung unterschiedliche Strategien erforderte.

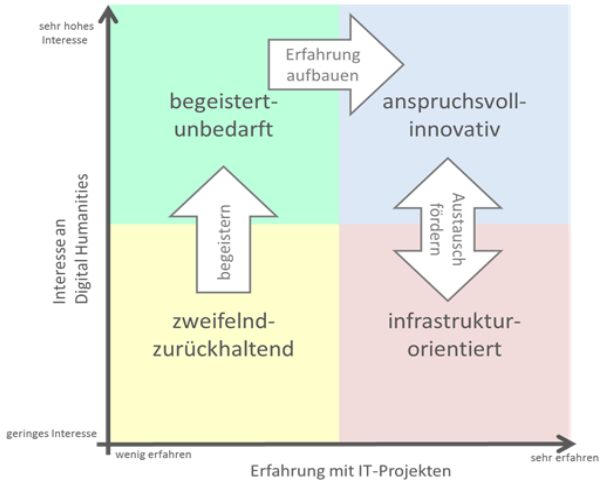

Abb. 2: Portfoliodarstellung der Nutzerschaft mit Strategien zu deren Einbindung

[9] Für das Projektmanagement wurden *Redmine und GanttProject* verwendet. Insbesondere durch die unterschiedlichen Organisationsstrukturen der Häuser, war ein gemeinsames Projektmanagement notwendig und erleichterte das Arbeiten im Verbundprojekt erheblich. Vgl. https://www.ganttproject.biz/ (Zugriffsdatum: 3.1.2019) und https://www.redmine.org/ (Zugriffsdatum: 3.1.2019). Zum Projektmanagement vgl. Reed, Ashley: Managing an established digital humanities project: Principles and practices from the twentieth year of the William Blake archive. In: DHQ, 8,1 (2014). http://www.digitalhumanities.org/dhq/vol/8/1/000174/000174.html (Zugriffsdatum: 3.1.2019).

zweifelnd-zurückhaltende Gruppe

- gekennzeichnet durch: geringes Interesse an Themen der *Digital Humanities* und keine/geringe Erfahrungen mit IT-Projekten.
- Beispiel: Es wird verneint, dass im Forschungsprozess Daten genutzt oder generiert werden.
- Strategie: Die Gruppe kann durch Ergebnispräsentation anderer Forscher/Innen, vor allem aus der begeistert-unbedarften Gruppe überzeugt werden. Die Personen sollten zu Ergebnispräsentationen eingeladen werden, der Kontakt soll aufrechterhalten werden.

begeistert-unbedarfte Gruppe

- gekennzeichnet durch: sehr hohes Interesse an Themen der *Digital Humanities* und geringe Erfahrungen mit IT-Projekten.
- Beispiel: Es werden Vorschläge für Analysen und Visualisierungen eingebracht, es liegen aber noch keine Daten/nicht in ausreichender Qualität vor.
- Strategie: Erfahrung durch kleine Projekte aufbauen, Ergebnisse präsentieren, Netzwerke aufbauen.

anspruchsvoll-innovativ Gruppe

- gekennzeichnet durch: sehr hohes Interesse an Themen *der Digital Humanities* und viel Erfahrung mit IT-Projekten.
- Beispiel-Anforderungen: Es werden Vorschläge eingebracht, die sehr viele Ressourcen in der Umsetzung verschlingen würden und nur für eine kleine Anzahl spezialisierter Forscher/Innen von Nutzen sein würde. Es werden viele Vorschläge eingebracht.
- Strategie: Früh in Pläne einbinden und um Feedback bitten. Für Verständnis werben, dass einige der sehr guten und innovativen Vorschläge nur schwer umzusetzen wären und neue Infrastruktur benötigten. Austausch mit infrastruktur-orientierten Gruppe fördern, um realistische Pläne zu konzipieren.

infrastruktur-orientierte Gruppe

- gekennzeichnet durch: geringes Interesse an Themen der *Digital Humanities* und viel Erfahrung mit IT-Projekten
- Beispiel-Anforderungen: Es werden realistische Anforderungen genannt, die jedoch wenig innovativ sind, die Umsetzung sollte sofort erfolgen.
- Strategie: Für Forschungsthemen begeistern, in Planungsprozesse einbeziehen und deutlich machen, dass der Aufbau neuer Infrastrukturen (z.B. digitale Langzeitarchivierung) Zeit braucht und zunächst mit der anspruchsvoll-innovativen Gruppe diskutiert wird.

3. Profilbildung durch Diskurs

Die in den ersten beiden Jahren gesammelten Anforderungen führten zu einer Profilschärfung der im Antrag nur skizzierten Bedarfe und Funktionen der digitalen Infrastruktur im Forschungsverbund MWW. Im Kontext der Profilbildung und der ersten Präsentation des Umsetzungskonzeptes entwickelte sich ein Diskurs, der in unterschiedlichen Ausprägungen kontinuierlich die Infrastrukturprojekte begleitete und sich auch in Teilen der vorgestellten Nutzergruppen wiederfindet. Es ging dabei um die Verortung von Infrastrukturaufbau zwischen Dienstleistung und Forschung.[10] Die Position für das Vorhandensein von Forschungsanteilen folgte der Argumentation, dass es sich um anwendungsorientierte Forschung handele, die ihren Ausdruck in der Notwendigkeit der vor der technischen Umsetzung zu erfolgenden konzeptionellen Arbeiten erfährt, die sich dezidiert auf die Modellierung und Abbildung von Forschungsprozessen im digitalen Raum fokussierte. Auf der anderen Seite wurde die Distanz zu etablierten geisteswissenschaftlichen Verfahren als Belege für die Einstufung von Infrastrukturaufbau als Dienstleistung angeführt.[11] An dieser Stelle soll nicht der Diskurs in allen Einzelheiten wiedergegeben werden, jedoch ist für den vorliegenden Beitrag relevant zu betonen, dass die Frage des Forschungsanteils bei Infrastrukturprojekten auch Auswirkungen auf die praktische Arbeit hatte.[12] Denn wenn Infrastrukturaufbau als Dienstleistung verstanden wird, gilt das Primat der Forschung und die Bedarfe der Forschung werden höher gewichtet als technische Aspekte bzw. Limitationen. Auch intendiert eine Dienstleistung immer die gelungene Umsetzung der gestellten Forderung. Begreift man Infrastrukturaufbau als Forschung kommt den Aspekten der experimentellen Auslotung von potentiellen Lösungen durch Mock-ups und Prototypen, der Beurteilung eines negativen Ergebnisses als Erfolg und der nach einer Eruierung erfolgten Ablehnung von nicht erfüllbaren Bedarfen eine ungleich höhere Bedeutung zu. Während diese Faktoren in der Informatik bekannt und etabliert sind, stoßen sie in den Geisteswissenschaften in der Regel nicht auf Akzeptanz. Es ist hier Aufgabe der *Digital Humanities* nicht nur ausgewählte Konzepte der Informatik zu adaptieren, sondern sie auch entsprechend zu kommunizieren - erfolgreiche Infrastruktur kann nur im Dialog zwischen den potentiellen

[10] Ein solcher Diskurs ist sicherlich kein Spezifikum der *Digital Humanities*, sondern kann in den Kontext interdisziplinärer Zusammenarbeit verortet werden, wenn Disziplinen mit unterschiedlicher Methodik und Fachvokabular zusammenkommen. Es kann aber diskutiert werden, ob aufgrund der größeren Ferne der Informatik von den Geisteswissenschaften hier nicht ein größeres Potential für Missverständnisse vorliegt. Dies im Besonderen weil digitale Methoden sich zunehmend zu einem festen Bestandteil innerhalb der Geisteswissenschaft entwickelt. Vgl. Busch, Hannah / Rettinghaus, Klaus / Schrade, Torsten / Schulte, Stefan: Aktuelle Situation der RSEs: Karriere - welche Karriere? In: Blog der Research Software Engineering in den Digital Humanities (DH-RSE). 20. August 2018. https://dh-rse.github.io/workshop/dhd2018/karriere/anerkennung/arbeitsbedingungen/2018/08/20/aktuelle-situation-der-rses-karriere-welche-karriere.html (Zugriffsdatum: 3.1.2019).

[11] Zu Verfahren in den DH vgl. Reiche, Ruth / Becker, Rainer / Bender, Michael / Munson, Matthew / Schmunk, Stefan / Schöch, Christof: "Verfahren der Digital Humanities in den Geistes- und Kulturwissenschaften" (DARIAH-DE Working Papers, 4). Göttingen 2014. urn:nbn:de:gbv:7-dariah-2014-2-6 (Zugriffsdatum: 3.1.2019).

[12] Auf dem MWW-Barcamp Data and Demons: Von Bestandsdaten zu Services zeigte sich, dass viele DH-Projekte von diesem Diskurs betroffen sind, der aber bisher wenig Eingang in eine öffentliche Diskussion gefunden hat. So wurden folgende Themen diskutiert: “Dienstleistung, Forschung, eierlegende Wollmilchsau”, “Was ist eigentlich eine DH-Fragestellung?” oder “Verhältnis DH / Hermeneutik.” Vgl. Dogunke, Swantje / Steyer, Timo / Mayer, Corinna: Barcamp Data and Demons: von Bestands- und Forschungsdaten zu Services. Treffen sich ein Bibliothekar, eine Archäologin, ein Informatiker, … . In: LIBREAS. Library Ideas, 33 (2018). https://libreas.eu/ausgabe33/dogunke/ (Zugriffsdatum: 3.1.2019).

Nutzern/Innen und den für den Aufbau und Unterhalt der Infrastruktur verantwortlichen Personen erfolgreich gestaltet werden.[13]

Gleiches gilt für die generierten Forschungsdaten, hier müssen die inhaltliche und technische Expertise kombiniert werden, um ein qualitativ hochwertiges Datenset zu erzeugen, welches für computergestütze Analyse ebenso geeignet ist wie für die Nachnutzung durch andere Forschungsvorhaben. Die Herausforderungen und Aufgaben der *Data Stewardchip* können nur in multidisziplinären Teams mit technischer und inhaltlicher Expertise und unter Beachtung aktueller Richtlinien und Anwendungsprofile gemeistert werden.

Die Antwort auf den Forschungsanteil an Infrastruktur kann dabei nicht pauschal beantwortet werden, zu heterogen sind die Aufgaben. Die Deklaration bestimmter Elemente oder Prozesse als Forschung sollte durch die Fachdisziplin definiert werden, welche die größere inhaltliche Nähe zu den Vorhaben aufweist. Auch innerhalb der *Digital Humanities* würde z. B. die Frage, ob die Entwicklung eines Metadatenformates Forschungsanteile aufweist, auf geteilte Meinungen stoßen. Zudem ist die fachliche Zuordnung der einzelnen Komponenten der digitalen Infrastruktur im MWW sehr unterschiedlich, auch wenn alle unter den Begriff der *Digital Humantities* zusammengefasst werden. Der Aufbau des Forschungsraum ist mehr durch Forschungsprozesse im Bereich der *Digital Humanities* definiert, der Verlässlichen Speicher hat größere informatische Elemente während die bestandsübergreifende Suche im Bereich der Informationswissenschaft *(Information Retrieval*) zu verorten ist.[14] Hinzu kommt, dass eine Trennung von *Digital Humanities* und IT-Dienstleistungen nicht immer stringent möglich ist.
Es ist sicherlich ratsam zu betonen, dass nur durch eine entsprechende digitale Infrastruktur bestimmte Forschungsfragen gestellt werden können - Infrastrukturaufbau dient daher der Ermöglichung von Forschung im digitalen Bereich und ist somit zumindest eng mit der Forschung verbunden. Zwar ist der Diskurs über den Forschungsanteil von Infrastrukturprojekten noch nicht abschließend geklärt und solange die Kanonisierung der *Digital Humanities* nicht abgeschlossen ist, kann auch keine endgültige Lösung erfolgen. Auch spiegelt sich hier die Diskussion um die wissenschaftliche Verortung von Grundlagenforschung wider.

Für die Entwicklung des virtuellen Forschungsraums brachte dieser Diskurs aber auch wertvolle Impulse für die Profilbildung, die sich neben der Konzentration auf den Bereich der Sammlungsforschung vor allem in der Notwendigkeit einer kontinuierlichen Wissenschaftskommunikation mit den Nutzer/Innen auswirken.[15] Dies trifft ebenso auf die Entscheidung zu, den Forschungsraum auf wenige Tools zu konzentrieren, diese aber besser hinsichtlich ihrer Potentiale und Grenzen zu dokumentieren und zu kuratieren, so dass realistische Anwendungsfälle generiert werden. Ein wichtiger Teil der Profilbildung besteht in der Abgrenzung zu anderen Infrastrukturen in einem positiven Sinne, so wurde bewusst der Bereich der digitalen Editionen nicht als eine Komponente des Virtuellen Forschungsraum identifiziert, weil

[13] Das Problem scheint auch darin zu bestehen, dass es wenig gemeinsamen Austausch gibt, die *Digital Humanities* haben mittlerweile eigene Publikations- und Kommunikationsstrukturen aufgebaut. So spielen wissenschaftliche Blogs oder Twitter eine größere Rolle als bei den "traditionellen" Geisteswissenschaften.

[14] Die unterschiedliche formale Abteilungszuordnung der DH-Mitarbeiter in den Institutionen spiegelt hier vielleicht auch das Rollenverständnis und die Bewertung des Forschungsanteils wider.

[15] Hier erwies sich die Taxonomie TaDIRAH (Taxonomy of Digital Research Activities in the Humanities) als gemeinsames Vokabular zu Forschungsprozessen als hilfreich. Die Erläuterungen sind für Informatiker/Innen, DH-Mitarbeiter/Innen und Geisteswissenschaftler verständlich.

bereits u. a. mit *Textgrid* eine umfassende Infrastruktur für diesen Bereich existiert.[16] Der Zuschnitt des Forschungsraums orientierte sich an den vorhandenen Sammlungen der Häuser und den nicht durch andere Infrastrukturen abgedeckten aber auch realistisch umzusetzenden Bedarfen der Forschungsprojekte.

4. Von der Konzeption zur Umsetzung

Die Analyse der Nutzerschaft, der vorhandenen Datenbestände und der in den beteiligten Institutionen vorhandenen Infrastruktur lieferte wichtige Erkenntnisse für die genauere Konzeption des Forschungsraums. Als infrastrukturelle Basis des Forschungsraums wurde die Notwendigkeit eines Content Management System identifiziert. Das CMS dient als zentrales System, über welches die weiteren Komponenten zusammengeführt sowie die Interaktion mit den Nutzern/Innen und der Community erfolgt. Hierfür wurde sich für das Content Management System *Liferay* entschieden, welches für die beschriebene Nutzung von der Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen (GWDG) für MWW aufgesetzt und modifiziert wurde.[17] Während die typischen Funktionalitäten eines CMS bereits sehr früh zum Einsatz kamen, wie z. B. in der projektinternen Kommunikation und Verwaltung aber auch in der Planung und Durchführung von Veranstaltungen[18], wurden die anderen Komponenten (Suche, Tools) weitestgehend unabhängig vom CMS entwickelt und erst in einer funktionalen Version mit dem CMS über die Entwicklung eigener Portlets verbunden.

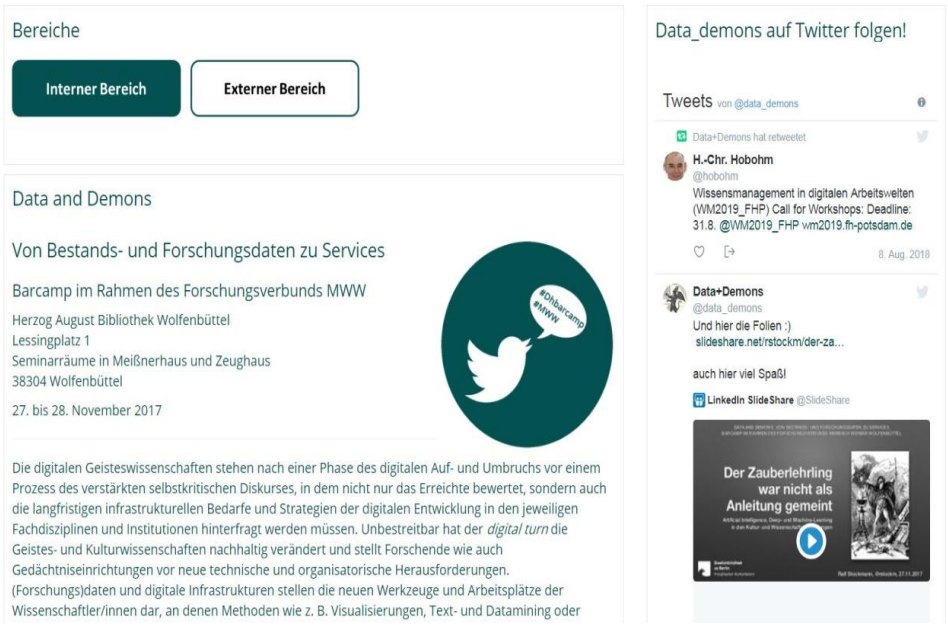

Abb.3: Ankündigung des Barcamps "Data and Demons" im virtuellen Forschungsraum. Auf der rechten erscheint die Twitter-Timeline zur Veranstaltung

[16] vgl. Neuroth, Heike / Rapp, Andrea / Söring, Sibylle (Hrsg.): TextGrid: Von der Community — für die Community. Eine Virtuelle Forschungsumgebung für die Geisteswissenschaften. Glückstadt: Verlag Werner Hülsbusch 2015.
[17] Ein Prototyp kann eingesehen werden unter: https://vfr.mww-forschung.de/ (Zugriffsdatum: 3.1.2019).
[18] So wurde der virtuelle Forschungsraum für die Organisation und Kommunikation des Barcamps "Data and Demons: Von Bestandsdaten zu Services" verwendet. Vgl. https://vfr.mww-forschung.de/web/barcamp-data-and-demons (Zugriffsdatum: 3.1.2019).

Gerade die aus den Nutzerinterviews hervorgegangenen Wünsche für die durch den Forschungsraum unterstützten digitalen Methoden sowie die dafür benötigten Services und Tools stellten dabei eine besondere Herausforderung dar. Die Tools und Services wurden zunächst ohne konkrete technische Umsetzungsszenarien beschrieben und unter der Verwendung der Taxonomie TaDiRAH (Taxonomy of Digital Research Activities in the Humanities) kategorisiert.[19]

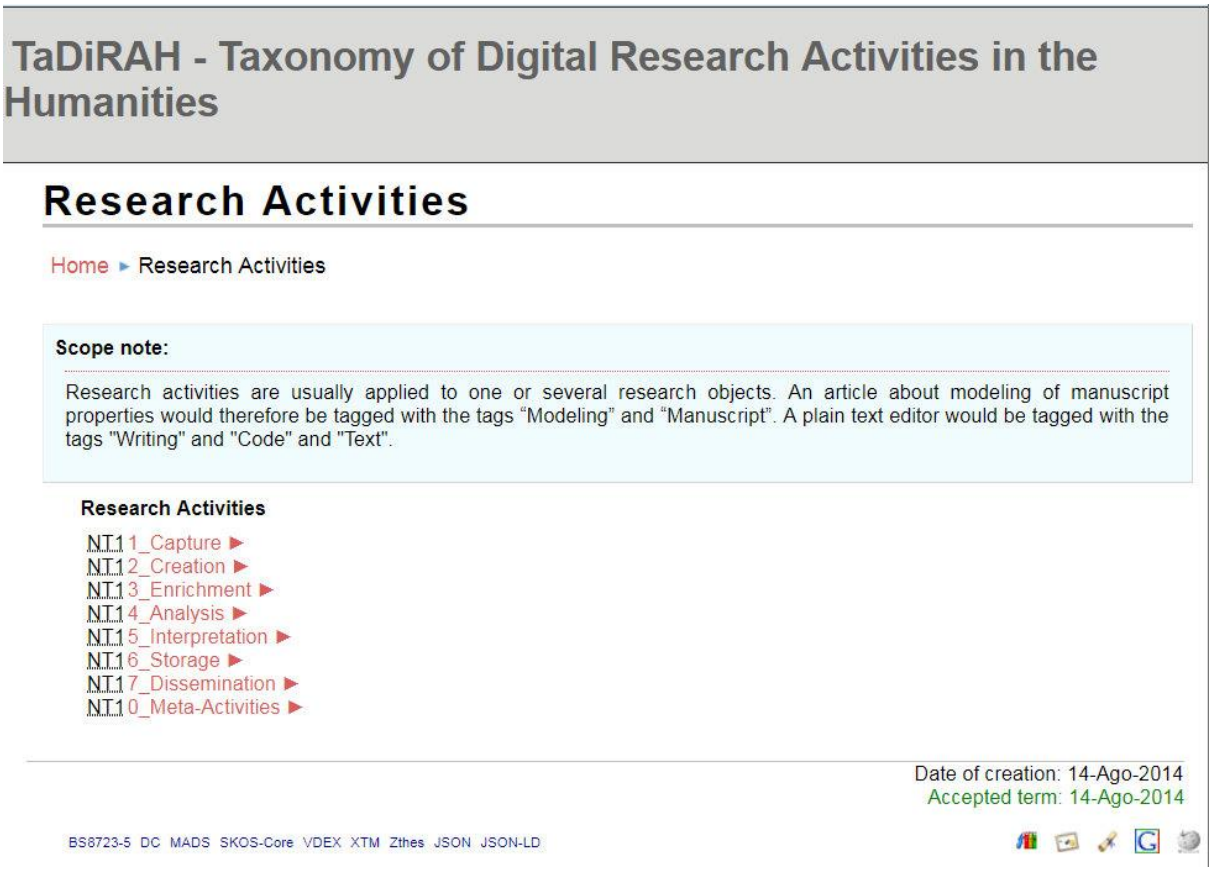

Abb. 4: Screenshot von TaDiRAH: http://tadirah.dariah.eu/vocab/index.php?tema=6

Erst im nächsten Schritt wurde eine Marktsicht über bestehende Tools vorgenommen, die für die evaluierten Verwendungszwecke generell in Frage kamen. Es stellte sich dabei heraus, dass für viele Methoden mehrere Tools bereits existierten, welche aber den befragten Wissenschaftlern/Innen nicht bekannt waren.[20] Dabei konnten zwei Arten von Tools unterschieden werden: Zum einen basale Tools für alltägliche Arbeiten und zum anderen DH-Tools für die Arbeit mit den Daten. Die basalen Tools erleichtern und unterstützen den Forschenden in der Kommunikation und im kollaborativen Arbeiten, wie z. B. im Austausch von Dateien, in

[19] Unter Services werden Angebote verstanden, welche automatisiert im Hintergrund ausgeführt und bei denen der Nutzer wenig bis keine direkte Interaktion vornehmen muss. Tools dagegen verlangen eine direkte Nutzerinteraktion und dementsprechend auch häufig eine Einarbeitung des Nutzers in die Funktionsweise des jeweiligen Tools.
Zu TaDiRAH vgl. http://tadirah.dariah.eu/vocab/index.php (Zugriffsdatum: 3.1.2019) und Borek, Luise: TaDiRAH – Taxonomy of Digital Research Activities in the Humanities. In: DHd-Blog. 19. Februar 2014. https://dhd-blog.org/?p=3073 (Zugriffsdatum: 3.1.2019).

[20] Einen Eindruck über die Vielzahl der bereits existierenden Tools liefert das Directory of Research Tools (DiRT): https://dirtdirectory.org/ (Zugriffsdatum: 3.1.2019).

der gemeinsamen Textproduktion oder in der Anlage von Bibliographien. Es sind globale Anwendungen, die jedes Projekt benötigt und für welche die Wissenschaftler/innen häufig proprietäre Produkte verwenden. Hier besteht der hauptsächliche Nutzen des Forschungsraums darin, den Wissenschaftlern/innen sichere und nicht von einem kommerziellen Anbieter abhängige Alternativen anzubieten:

- Eine Owncloudinstanz für die Verwaltung und das Teilen von Dateien löste bisherige Produkte wie Dropbox oder WeTransfer ab.
- Für die Literaturverwaltung in den Forschergruppen wurde Zotero ausgewählt und ersetzt Citavi und Endnote.
- Gemeinsame Texte wurden zuvor durch GoogleDocs oder komplizierte Umlaufverfahren mit Emailverkehr erstellt, nun bietet der Virtuelle Forschungsraum Etherpads und Onlyoffice als integrierte Lösungen an.

Schwieriger als die basalen Tools gestalte sich die Umsetzung der auf die spezielle Anwendung von DH-Methoden fokussierten Tools. Zwar existieren auch hier bereits für viele der Methoden entsprechende Tools, aber bei einer genaueren Evaluierung der Tools stellte sich heraus, dass zwar der digitale Wandel einen kontinuierlichen Strom an Tools generiert, allerdings die langfristige Wartung und Aktualisierung der Tools keine Selbstverständlichkeit darstellen. Hinzu kommt, dass die Toolentwicklung häufig im Kontext von Drittmittelprojekten stattfindet, was dazu führt, dass die entstehenden Tools auf ihre Verwendung im Projekt fokussiert werden. Zwar ist ihre Nachnutzung durch die Forschung stets intendiert, aber eine entsprechende Unterstützung, z. B. durch eine leicht zu parametrisierende Konfiguration oder eine Dokumentation, findet selten statt. Für eine communitygetriebene Weiterentwicklung fehlt es meistens an der kritischen Masse und/oder an Wissen sowie personellen Kapazitäten.

Die Auswahl der Tools für den Forschungsraum erfolgt nach einem Bewertungssystem, welches folgende Faktoren umfasste:

- Eignung für die identifizierte Methode
- Bevorzugung von Open Source Produkten
- Existenz von Dokumentation, Ansprechpartnern oder einer Community
- Referenzen für den Einsatz des Tools
- Technische Realisierung
- Aufwand der Implementierbarkeit in den Virtuellen Forschungsraum
- Austauschmöglichkeiten mit anderen Tools

Entsprechend der Ergebnisse konnten die Tools in drei Entwicklungsstufen eingeordnet werden:

- Tool existiert und kann ohne größere Anpassung integriert werden (z. B. Etherpad, Zotero[21])

[21] vgl. Puckett, Jason: Zotero: A Guide for Librarians, Researchers and Educators. Zweite Auflage. Chicago: AMER LIB ASSN 2017.

- Tools existiert, muss aber für die Verwendung noch anpasst werden bzw. das Wissen für die Integration war bei MWW nicht vorhanden (z. B. *VIKUS Viewer*[22], *OMEKA*)
- Tool existiert noch nicht und muss entwickelt werden (z. B. *LibReTo*)

Insbesondere die Tools der zweiten Kategorie bedingten den Aufbau von Kooperationen mit den Entwicklern oder den sogenannten Power Usern der Tools. Die Nachnutzung des jeweiligen Tools im MWW-Kontext, wurde stets begrüßt und teilweise sogar mit Ressourcen unterstützt. So entstanden schnell Entwicklungsgemeinschaften, die über den unmittelbaren Zweck auch Austausch und Diskussion ermöglichen.

Während der ersten Förderphase von MWW entstanden durch die beteiligten Forschungsprojekte auch neue Toolvorschläge bzw. Modifikationswünsche an integrierten Tools. So wurde im Projekt Bildpolitik die virtuelle Ausstellung *Luthermania* mit dem Open Source Framework *OMEKA* realisiert. Das Portfolio des Forschungsraums ist daher dynamisch ausgelegt, es kann entsprechend den Rückmeldungen der Community oder den sich aus den Forschungsprojekten ergebenden Bedarfen erweitert und modifiziert werden. Auch die Entfernung von nicht (mehr) genutzten Tools ist eine Option, um den Pflege- und Wartungsaufwand anzupassen. Ein entscheidender Faktor bei allen Angeboten ist aber das Vorhandensein von Kompetenz für die Tools und Ansprechpartner für die Projekte - die Forschung benötigt kuratierte Tools, um auch ein Vertrauen für ihre nachhaltige Nutzung zu sichern.

Abgesehen von den Tools wurden über das CMS noch projektspezifische Bereiche eingerichtet, in denen die Informationen, Bibliographien und Dateien des Projektes verwaltet und ausgetauscht werden konnten. Einzelne Projekte publizierten auch Projektergebnisse über das CMS, so entsteht aus dem Projekte Text und Rahmen ein Portal zu mittelalterlichen Psaltern. Andere Projekte verwendeten die existierenden Infrastrukturen der Häuser, auch Kombinationen sind möglich. Aus dem soeben erwähnten Projekt entstand auch eine Edition eines niederdeutschen Psalters, welches in der Wolfenbütteler Digitalen Bibliothek publiziert worden ist.[23] Als wichtiger Faktor erwies sich stets das Vorhandensein von Schnittstellen, um Daten automatisiert auszutauschen und keine redundante Datenhaltung zu erzeugen. Es wird auch in Zukunft keine globale Infrastruktur für alle Bedürfnisse der digitalen Geisteswissenschaften geben, sondern es ist von einem Netzwerk an Infrastrukturen auszugehen, deren gegenseitige Kommunikation auf Ebene der Daten nur über Schnittstellen und Kooperationen zu gewährleisten ist. Inwieweit sich hier die Nationale Forschungsdateninfrastruktur zukünftig verorten wird, ist eine wegweisende Frage.[24]

[22] vgl. Moeller, Katrin: Vikus: Ein Visualisierungstool für Korpus- und Massendaten. In: Digitale Geschichtswissenschaft. Das Blog der AG Digitale Geschichtswissenschaft im VHD. 15. Februar 2017. https://digigw.hypotheses.org/1577 (Zugriffsdatum: 3.1.2019).

[23] vgl. Lateinisch-mittelniederdeutscher Psalter mit Kommentar Cod. Guelf. 81.10 Aug. 2°. Ursula Kundert (Hg.), unter Mitarbeit von Hanne Grießmann. Diplomatische Teiledition. Wolfenbüttel 2017. http://diglib.hab.de/mss/81-10-aug-2f/start.htm (Zugriffsdatum: 3.1.2019).

[24] vgl. Positionspapier der Union der deutschen Akademien der Wissenschaften zur Schaffung einer Nationalen Forschungsdateninfrastruktur (NFDI) vom 5. Februar 2018. https://www.akademienunion.de/fileadmin/redaktion/user_upload/Publikationen/Positionspapiere/NFDI-Positionspapier_Akademienunion_01.pdf (Zugriffsdatum: 3.1.2019).

5. Zentraler Zugang zu digitalen Sammlungen

Neben den aus den MWW-Forschungsprojekten neu generierten (Forschungs-)daten rekrutiert sich der Datenbestand des Virtuellen Forschungsraums maßgeblich aus den bestehenden digitalen Sammlungen der beteiligten Einrichtungen. Im Gegensatz zu anderen Infrastrukturen kann der virtuelle Forschungsraum MWW daher bereits ab ovo eine umfangreiche Datenbasis zur Verfügung stellen, welche sukzessive erweitert wird.[25] Der Synergieeffekt durch die fast gleichzeitige Integration von Daten und Tools in einer Infrastruktur besteht in der Möglichkeit, beide Bestandteile hinsichtlich unterstützter Formate und Anwendungsprofile von Meta- und Normdaten aufeinander abzustimmen.

Die Integration von existierenden Sammlungen in eine gemeinsame Infrastruktur birgt jedoch besondere Anforderungen im Bereich der Datenmodellierung und des Datenmanagements. Die für den virtuellen Forschungsraum verfügbaren Sammlungen bestehen vor allem aus Metadaten und Digitalisaten, welche analoge Bestände beschreiben und erschließen. Die Erstellung der Datensets der über 50 digitalen Sammlungen erfolgte in der Regel über Drittmittelprojekte, bei denen jeweils eigene Datenmodelle und technische Realisierungen entwickelt und umgesetzt worden sind.[26] Viele dieser Projekte sind bereits seit Jahren abgeschlossen und es findet keine aktive Datenerhebung mehr statt.[27]

Als Beispiele zu nennen wären hier u.a.:

- *Forschungsportal zur Wissensproduktion an der Universität Helmstedt*
- *Briefrepertorium zu Johann Wolfgang von Goethe*
- *Katalog der Leichenpredigten der Leichenpredigten*
- *Mitgliederverzeichnis der Fruchtbringenden Gesellschaft*

Durch die Zusammenführung und Aufbereitung der Sammlungen im virtuellen Forschungsraum bzw. in der Metadatengrammatik soll zum einen eine interoperable Datenbasis - in Orientierung an den FAIR-Date-Prinzipien für die wissenschaftliche Nutzung der Daten - mit den im Forschungsraum angebotenen Tools und Services zur Verfügung gestellt werden.[28] Zudem soll durch den Zusammenschluss der Sammlungen eruiert werden, welche gemeinsame Standards und Erschließungsrichtlinien für zukünftige digitale Erschließungsprojekte erforderlich sind und als Best-Practice-Beispiele für Sammlungserschließung dokumentiert werden. Gegenwärtig verhindern die unterschiedlichen technischen Systeme, Datenmodelle und Formate eine gemeinsame Nutzung der digitalen Sammlungen. Auch sind einige der Sammlungen bereits vor über zehn Jahren entstanden - während die Daten nichts von ihrer wissenschaftlichen Relevanz eingebüßt haben, entsprechen die Interfaces und Datenmodelle aber nicht mehr den aktuellen Standards und Gepflogenheiten. Im virtuellen Forschungsraum werden

[25] Aktuell sind circa 250.000 Dokumente im virtuellen Forschungsraum integriert.
[26] Eine Übersicht über die Sammlungen ist zu finden unter: https://vfr.mww-forschung.de/die-sammlungen (Zugriffsdatum: 3.1.2019).
[27] Selbstverständlich werden Korrekturen oder Ergänzungen vorgenommen, wenn neue Erkenntnisse vorliegen, die Anzahl der Datensätze wird jedoch vermutlich unverändert bleiben.
[28] Vgl. Kraft, Angelina: The FAIR Data Principles for Research Data. In: TIB-Blog – Weblog der Technischen Informationsbibliothek (TIB). 12. September 2017. https://blogs.tib.eu/wp/tib/2017/09/12/the-fair-data-principles-for-research-data/ (Zugriffsdatum: 3.1.2019).

die Sammlungen mittels eines gemeinsamen Interfaces zugänglich gemacht und so ein zentralen Zugang zu den heterogenen Sammlungen geschaffen. Um diese Ziele technisch zu realisieren, wird eine MWW-spezifische Instanz der DARIAH-DE Datenföderationsarchitektur verwendet.[29]

Eine zentrale Komponente der Datenföderationsarchitektur ist die Beschreibung der MWW-Sammlungen in der DARIAH-De Collection Registry hinsichtlich ihrer inhaltlichen, administrativen und technischen Merkmale.[30] MWW verwendet über Schnittstellen diese Beschreibungen auch für eine Sammlungsübersicht im Forschungsraum.

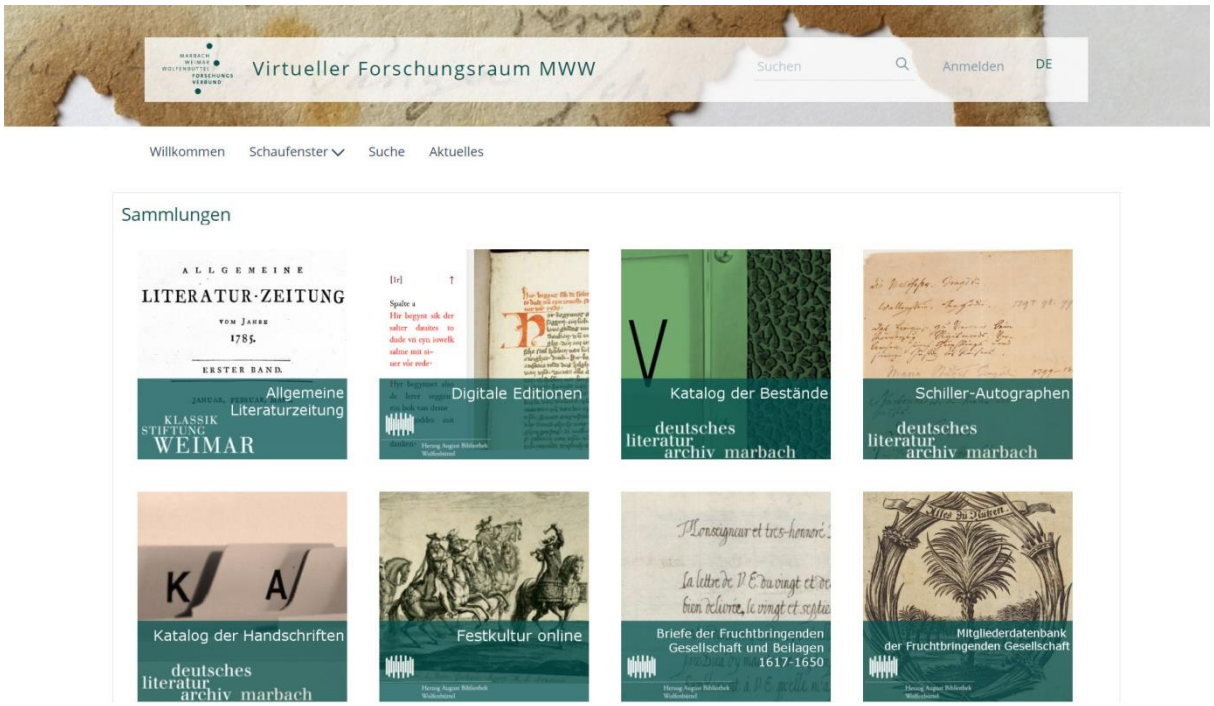

Abb. 5: Sammlungsübersicht im virtuellen Forschungsraum MWW. https://vfr.mww-forschung.de/die-sammlungen

Aufbauend auf den Sammlungsbeschreibungen werden die Datenmodelle der Sammlungen in der Data Modelling Environment angereicht und gemappt. Das im Projekt Datenmodellierung und Metadaten entwickelte gemeinsame Metadatenformat des Verbundes (MMM: MWW-Metadada-Model) bildet die Basis für die Harmonisierung der Sammlungen auf ein gemeinsames Metadatenschema. Alle Sammlungen werden daher auf MMM gemappt. Zwar können stark domänenspezifische Informationen aus den einzelnen Sammlungen nicht im gemeinsamen Metadatenschema abgebildet werden, nichtsdestotrotz stellt das Metadatenschema die Basis für die sammlungsübergreifende Suche (Generic Search) dar, in welcher die Sammlungen in ihren Kerninformationen durchsucht werden können.[31] Dafür bedarf es einer Anreicherung und Standardisierung der Sammlungen, da entweder erforderlichen Informationen für eine sammlungsübergreifende Suche in den Originaldaten fehlen (z. B. Datenlieferant, Link

[29] https://de.dariah.eu/data-federation-architecture (Zugriffsdatum: 3.1.2019).
[30] https://colreg.mww-forschung.de/colreg/ (Zugriffsdatum: 3.1.2019).
[31] https://search.mww-forschung.de/search/ (Zugriffsdatum: 3.1.2019).

zum Webauftritt) oder die Eingabe der Daten nicht dem gemeinsamen Standard des Metadatenformates entspricht (Datumsangabe nach ISO 8601, Sprachangaben nach ISO 639-2).

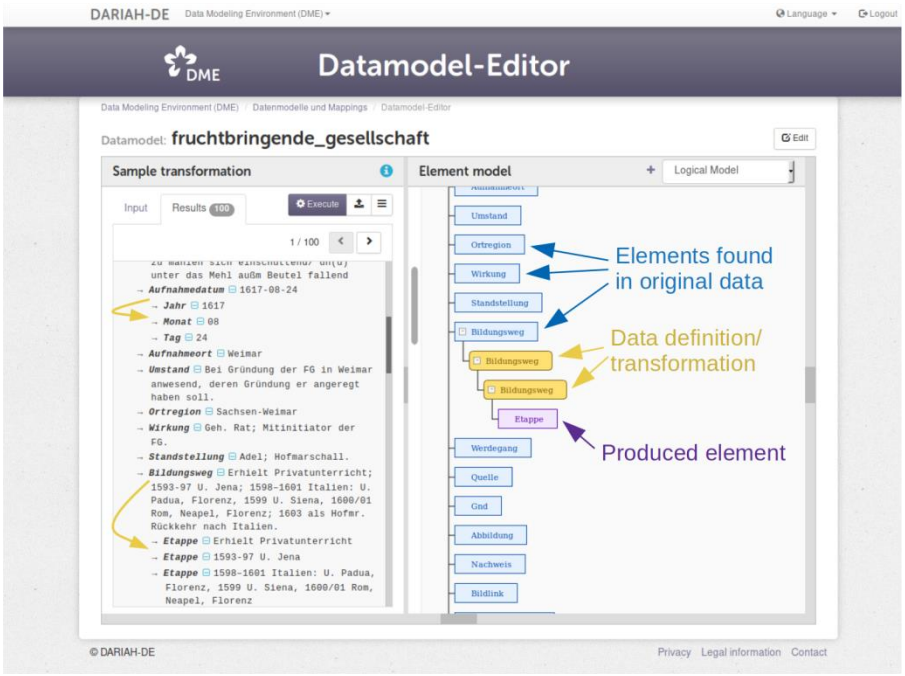

Abb. 6: Beispiel einer Datenanreicherung für die Sammlung "Mitgliederverzeichnis der Fruchtbringenden Gesellschaft

Den Weg der Abbildung von Sammlungen auf ein globales Modell verfolgen auch Infrastrukturen wie die *Europeana* und die *Deutsche Digitale Bibliothek*. Der Nachteil bei dieser Herangehensweise ist der Verlust von Informationen, die durch das Mapping auf den gemeinsamen Standard verloren gehen.[32] Um diesen Nachteil auszugleichen, sind die integrierten Sammlungen in der sammlungsübergreifenden Suche auch in ihrer vollen Komplexität durchsuchbar, dann jedoch nur innerhalb des von ihnen verwendeten spezifischen Datenmodells. Dadurch können die Sammlungen sowohl aus einer globalen Perspektive wie auch aus einer domänenspezifischen Sicht recherchiert werden. Der Nachteil dieser Vorgehensweise liegt in der Komplexität der Bedienung, denn während die globale Suche einen zentralen Suchschlitz als Einstieg bietet, müssen bei den domänenspezifischen Suche die Datenmodelle und die entsprechenden Elemente ausgewählt werden. Hier bedarf es Expertenwissens oder zumindest einer umfangreichen Einarbeitungszeit. Aber die Frage nach der benötigten Komplexität im Verhältnis zur Erwartung der Nutzer/Innen stellt eine immer wiederkehrende Frage dar, die beim Aufbau einer jeden digitalen Infrastruktur stets vom neuen beantwortet werden muss.

[32] vgl. Gradl, Tobias / Henrich, Andreas / Plutte, Christoph: Heterogene Daten in den Digital Humanities: Eine Architektur zur forschungsorientierten Föderation von Kollektionen. In: Constanze Baum und Thomas Stäcker (Hg.): Grenzen und Möglichkeiten der Digital Humanities (Sonderband der ZfdG, 1). 2015. DOI: 10.17175/sb001_020 (Zugriffsdatum: 3.1.2019).

6. Resümee

In der Leistungsabgrenzung des Vorhabens muss zwischen dem Aufbau und der Entwicklung einer Infrastruktur - organisierst als zeitlich begrenztes Projekt - und dem Betrieb dieser Infrastruktur unterschieden werden. Der Betrieb von digitaler Infrastruktur ist stets mit dem Wunsch der Dauerhaftigkeit und Nachhaltigkeit verbunden - diese Faktoren sind auch entscheidend für die Bildung einer Community und die wissenschaftliche Akzeptanz. Der virtuelle Forschungsraum MWW hat den Vorteil der Bindung an die beteiligten Häuser und an die dortigen Forschungsprojekte, so dass hier ein signifikanter Anteil der angesprochenen Community bereits impulsgebend in die Entwicklung eingebunden werden konnte. Auch erzeugt die institutionelle Anbindung Vertrauen bei den Nutzer/Innen, da die Einrichtungen auch über MWW hinaus den Unterhalt der aufgebauten Infrastruktur garantieren. Der virtuelle Forschungsraum vereint daher den Vorteil einer lokalen Infrastruktur, welche in den etablierten Workflows und personellen Strukturen der Häuser eingebunden ist mit dem Zuschnitt einer fachspezifischen Infrastruktur. Auch wenn einzelne Komponenten, wie die sammlungsübergreifende Suche, generischen Charakter haben, fokussiert sich der Forschungsraum auf die Sammlungsforschung. Dies auch vor dem Hintergrund, dass die Geisteswissenschaften aufgrund ihrer Diversität einen erhöhten Bedarf an fachspezifischen Lösungen haben und der Aufbau und Unterhalt von generischen Infrastrukturen Ressourcen verlangt, die der Verbund durch seine Projektstruktur nicht tragen könnte.

Der fachspezifische Zuschnitt bedingt einen regelmäßigen Abgleich der Angebote des Forschungsraums mit den Ansprüchen und Wünschen der Forschung, um eine stetige Weiterentwicklung und Anpassung des Forschungsraums zu gewährleisten. Die Erfahrung aus anderen Projekte lehrt, dass Forschende nicht darauf warten, dass Ihnen die benötigten Services oder Tools zur Verfügung gestellt werden, sondern dazu neigen auf verfügbare Alternative zurückzugreifen, selbst wenn diese nicht genuin für wissenschaftliches Arbeiten gedacht sind. Es muss daher das Ziel jeder digitalen Infrastruktur sein, in das Portfolio der Wissenschaftler/Innen aufgenommen zu werden. Dieser Prozess kann nur gelingen, wenn die Forschungseinrichtungen es schaffen, die aufgebaute Infrastruktur langfristig anzubieten und weiterzuentwickeln - eine Forderung die angesichts vieler veralteter Forschungsportale und Webseiten zu Forschungsprojekten keine Selbstverständlichkeit darstellt. Auf der anderen Seite müssen aber auch die Wissenschaftler/Innen Angebote wahrnehmen, die zwar die ihnen zugedacht Funktionen erfüllen, sich aber weder in ihrer Gestaltung noch im Support mit kommerziellen Angeboten messen können. Auch müssen die Rollen beim Aufbau von Infrastruktur ebenso klar definiert werden wie die zu erwartenden Erfolgskriterien. Die *Digital Humanities* sind dabei integrativer Bestandteil anzusehen, dessen Beteiligung bereits in der gemeinsamen Projektplanung und nicht erst in der Umsetzungsphase erfolgen sollte. Wird DH als integrativer Bestandteil der Forschungsprojektplanung verstanden, stehen benötigte Tools und Komponenten dann bereit, wenn sie benötigt werden. Zudem benötigen bestandsbezogene Forscher/Innen eine Legitimation durch einen Projektplan, um zeitintensive Prozesse wie Iterationen und Prototyping. Entscheidend für sie sind auch Belege für erfolgreiche DH-Projekte, um in den jeweiligen Fachdisziplinen ihre digitale Herangehensweise präsentieren zu können.

Ein weiterer wesentlicher Faktor ist die Akzeptanz von kollaborativen Arbeiten und Teamwork. Die korrekte Anwendung von Tools, die Generierung von qualitativ hochwertigen Forschungsdaten und die Durchführung von nachnutzbaren Forschungsergebnissen können nur

durch die Bündelung unterschiedlicher Kompetenzen gelingen. Daher kommen Kooperationen, externes Expertenwissen und der Nachnutzung bestehender Angebote entscheidende Bedeutung zu. Mindestens ebenso wichtig sind eine kontinuierliche Kommunikation mit den Nutzer/Innen und gemeinsame Veranstaltungen, um die Bedarfsermittlung weiterzuführen und gemeinsame Entwicklungsimpulse zu starten. Auch wenn der technische Kenntnisstand von Forschern/Innen zunehmend wächst, stellt die technische Entwicklung nicht nur eine Arbeitserleichterung dar, sondern bedeutet auch Mehrarbeit. Daher muss eine digitale Infrastruktur in bestehende Strukturen integriert bzw. vernetzt werden, um so den Forschungsprozess nicht nur technisch, sondern auch inhaltlich zu unterstützen. Eine digitale Infrastruktur wird nur dann von den Wissenschaftlern/Innen akzeptiert werden, wenn durch ihre Verwendung eine merkliche Arbeitsentlastung stattfindet und/oder der erzielte Mehrwert den erhöhten Aufwand rechtfertigt. Wenn der Aufbau von digitaler Infrastruktur sich auf das Anbieten von Daten und Tools beschränkt, wird nur eine temporäre Community aufgebaut, welcher sich aber nicht an die Infrastruktur binden lässt. Da die Ansprüche der Community - zumindest bei MWW- heterogen sind und auch einem wechselnden Bedingungen unterliegen, erscheint eine Mischung aus basalen, generischen mit domänenspezifischen Services und Tools ein erfolgversprechendes Konzept zu sein.

Daher gilt, was R. David Lankes zur Rolle von Bibliotheken im Kontext des digitalen Wandels gesagt hat, erst recht für digitale Infrastrukturen: „Bad Libraries Build Collections, Good Libraries Build Services, Great Libraries Build Communities".[33]

Literaturverzeichnis

Arbeitsgruppe Virtuelle Forschungsumgebungen in der Allianz der deutschen Wissenschaftsorganisationen (Hg.): Definition Virtuelle Forschungsumgebung, 2011. DOI: http://doi.org/10.2312/ALLIANZOA.028 (Zugriffsdatum: 3.1.2019).

Beyer, Hartmut / Münkner, Jörn / Schmidt, Katrin / Steyer, Timo: Bibliotheken im Buch: Die Erschließung von privaten Büchersammlungen der Frühneuzeit über Auktionskataloge. In: Hannah Busch / Franz Fischer / Patrick Sahle (Hg.): Kodikologie und Paläographie im digitalen Zeitalter 4 (Schriften des Instituts für Dokumentologie und Editorik, 11). Norderstedt: BoD 2017, S. 43-70.

Borek, Luise: TaDiRAH – Taxonomy of Digital Research Activities in the Humanities. In: DHd-Blog, 19. Februar 2014. https://dhd-blog.org/?p=3073 (Zugriffsdatum: 3.1.2019).

Busch, Hannah / Rettinghaus, Klaus / Schrade, Torsten / Schulte, Stefan: Aktuelle Situation der RSEs: Karriere - welche Karriere? In: Blog der Research Software Engineering in den Digital Humanities (DH-RSE), 20. August 2018. https://dh-rse.github.io/workshop/dhd2018/karriere/anerkennung/arbeitsbedingungen/2018/08/20/aktuelle-situation-der-rses-karriere-welche-karriere.html (Zugriffsdatum: 3. 1. 2019).

[33] vgl. Lankes, David R.: Beyond the Bullet Points: Bad Libraries Build Collections, Good Libraries Build Services, Great Libraries Build Communities (2012). https://davidlankes.org/beyond-the-bullet-points-bad-libraries-build-collections-good-libraries-build-services-great-libraries-build-communities/ (Zugriffsdatum: 3.1.2019).

Dogunke, Swantje / Steyer, Timo / Mayer, Corinna: Barcamp Data and Demons: von Bestands- und Forschungsdaten zu Services. Treffen sich ein Bibliothekar, eine Archäologin, ein Informatiker, … . In: LIBREAS. Library Ideas, 33 (2018). https://libreas.eu/ausgabe33/dogunke/ (Zugriffsdatum: 3.1.2019).

Gradl, Tobias / Henrich, Andreas / Plutte, Christoph: Heterogene Daten in den Digital Humanities: Eine Architektur zur forschungsorientierten Föderation von Kollektionen. In: Constanze Baum und Thomas Stäcker (Hg.): Grenzen und Möglichkeiten der Digital Humanities (Sonderband der ZfdG, 1). 2015. DOI: 10.17175/sb001_020 (Zugriffsdatum: 3.1.2019).

Kraft, Angelina: The FAIR Data Principles for Research Data. In: TIB-Blog – Weblog der Technischen Informationsbibliothek (TIB). 12. September 2017. https://blogs.tib.eu/wp/tib/2017/09/12/the-fair-data-principles-for-research-data/ (Zugriffsdatum: 3.1.2019).

Kramer, Michael J.: Going Meta on Metadata. In: Journal of Digital Humanities, 3, 2 (2014). http://journalofdigitalhumanities.org/3-2/going-meta-on-metadata/ (Zugriffsdatum: 3.1.2019).

Lankes, David R.: Beyond the Bullet Points: Bad Libraries Build Collections, Good Libraries Build Services, Great Libraries Build Communities (2012). https://davidlankes.org/beyond-the-bullet-points-bad-libraries-build-collections-good-libraries-build-services-great-libraries-build-communities/ (Zugriffsdatum: 3.1.2019).

Lateinisch-mittelniederdeutscher Psalter mit Kommentar Cod. Guelf. 81.10 Aug. 2°. Ursula Kundert (Hg.), unter Mitarbeit von Hanne Grießmann. Diplomatische Teiledition. Wolfenbüttel 2017. http://diglib.hab.de/mss/81-10-aug-2f/start.htm (Zugriffsdatum: 3.1.2019).

Moeller, Katrin: Vikus: Ein Visualisierungstool für Korpus- und Massendaten. In: Digitale Geschichtswissenschaft. Das Blog der AG Digitale Geschichtswissenschaft im VHD. 15. Februar 2017. https://digigw.hypotheses.org/1577 (Zugriffsdatum: 3.1.2019).

Neuroth, Heike / Rapp, Andrea / Söring, Sibylle (Hg.): TextGrid: Von der Community — für die Community. Eine Virtuelle Forschungsumgebung für die Geisteswissenschaften. Glückstadt: Verlag Werner Hülsbusch 2015.

Positionspapier der Union der deutschen Akademien der Wissenschaften zur Schaffung einer Nationalen Forschungsdateninfrastruktur (NFDI) vom 5. Februar 2018. https://www.akademienunion.de/fileadmin/redaktion/user_upload/Publikationen/Positionspapiere/NFDI-Positionspapier_Akademienunion_01.pdf (Zugriffsdatum: 3.1.2019).

Puckett, Jason: Zotero: A Guide for Librarians, Researchers and Educators. Zweite Auflage. Chicago: AMER LIB ASSN 2017.

Reed, Ashley: Managing an established digital humanities project: Principles and practices from the twentieth year of the William Blake archive. In: DHQ, 8,1 (2014). http://www.digitalhumanities.org/dhq/vol/8/1/000174/000174.html (Zugriffsdatum: 3.1.2019).

Reiche, Ruth / Becker, Rainer / Bender, Michael / Munson, Matthew / Schmunk, Stefan / Schöch, Christof: "Verfahren der Digital Humanities in den Geistes- und Kulturwissenschaften" (DARIAH-DE Working Papers, 4). Göttingen 2014. urn:nbn:de:gbv:7-dariah-2014-2-6 (Zugriffsdatum: 3.1.2019).

Stäcker, Thomas: Noch einmal: Was sind geisteswissenschaftliche Forschungsdaten? In: DHd-Blog. 6. Dezember 2015. https://dhd-blog.org/?p=5995 (Zugriffsdatum: 3.1.2019).

The New Media Consortium (Hg.): NMC Horizon Report. 2017 Library Edition. http://cdn.nmc.org/media/2017-nmc-horizon-report-library-EN.pdf (Zugriffsdatum: 3.1.2019).

Abbildungsverzeichnis

Abbildung 1: Erste Architekturskizze der Virtuellen Forschungsumgebung MWW @Eigene Grafik, 2018.

Abbildung 1: Portfoliodarstellung der Nutzerschaft mit Strategien zu deren Einbindung @Eigene Grafik, 2018.

Abbildung 2: Ankündigung des Barcamps “Data and Demons” im virtuellen Forschungsraum. Auf der rechten erscheint die Twitter-Timeline zur Veranstaltung: https://vfr.mww-forschung.de/web/barcamp-data-and-demons @Eigenhändiger Screenshot, 2018.

Abbildung 3: Screenshot von TaDiRAH: http://tadirah.dariah.eu/vocab/index.php?tema=6 @Eigenhändiger Screenshot, 2018.

Abbildung 4: Sammlungsübersicht im virtuellen Forschungsraum MWW. https://vfr.mww-forschung.de/die-sammlungen @Eigenhändiger Screenshot, 2018.

Abbildung 5: Beispiel einer Datenanreicherung für die Sammlung “Mitgliederverzeichnis der Fruchtbringenden Gesellschaft” @Eigenhändiger Screenshot, 2018.