

Heike Neuroth, Norbert Lossau, Andrea Rapp (Hrsg.)

Evolution der Informationsinfrastruktur

Kooperation zwischen Bibliothek und Wissenschaft



Neuroth/Lossau/Rapp (Hrsg.) · Evolution der Informationsinfrastruktur

**Heike Neuroth, Norbert Lossau,
Andrea Rapp (Hrsg.)**

Evolution der Informationsinfrastruktur

**Kooperation zwischen
Bibliothek und Wissenschaft**

Evolution der Informationsinfrastruktur

hg. von Dr. Heike Neuroth, Prof. Dr. Norbert Lossau, Prof. Dr. Andrea Rapp

Erschienen im Rahmen des zehnjährigen Jubiläums der Abteilung Forschung & Entwicklung 2012.

<https://sites.google.com/site/10jahrefe/>

Kontakt:

neuroth@sub.uni-goettingen.de

c/o Niedersächsische Staats- und Universitätsbibliothek Göttingen,

Dr. Heike Neuroth, Forschung & Entwicklung, Papendiek 14, 37073 Göttingen

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet unter <http://d-nb.de> abrufbar.

Die Inhalte dieses Buches stehen auch als Onlineversion (DOI:

<http://resolver.sub.uni-goettingen.de/purl/?webdoc-39006>) sowie über den Göttinger Universitätskatalog (<http://www.sub.uni-goettingen.de>) zur Verfügung. Die digitale Version steht unter folgender Creative-Commons-Lizenz: „Attribution-Noncommercial-Share Alike 3.0 Unported“ <http://creativecommons.org/licenses/by-nc-sa/3.0/>



Einfache Nutzungsrechte liegen beim Verlag Werner Hülsbusch.

© Verlag Werner Hülsbusch, Glückstadt, 2013

vwh

Verlag Werner Hülsbusch
Fachverlag für Medientechnik und -wirtschaft

www.vwh-verlag.de

In Kooperation mit dem Universitätsverlag Göttingen.

Markenerklärung: Die in diesem Werk wiedergegebenen Gebrauchsnamen, Handelsnamen, Warenzeichen usw. können auch ohne besondere Kennzeichnung geschützte Marken sein und als solche den gesetzlichen Bestimmungen unterliegen.

Korrektur und Satz: Werner Hülsbusch

Redaktionelle Betreuung und Umschlaggestaltung: Stefan Buddenbohm
(auf Grundlage einer Grafik von Ralf Stockmann)

Druck und Bindung: SOWA Sp. z o. o., Warszawa

Printed in Poland · ISBN: 978-3-86488-043-8

Danksagung

Die Abteilung Forschung & Entwicklung ist bei der Umsetzung ihrer Forschungs- und Entwicklungsprojekte maßgeblich auf die Unterstützung verschiedener Forschungsförderer angewiesen.

Die Durchführung der Mehrzahl der in diesem Band vorgestellten Vorhaben wäre ohne die großzügige Unterstützung durch die jeweiligen Forschungsförderer unmöglich gewesen. Diese Tatsache möchten wir zum Anlass nehmen, uns ausdrücklich bei allen Förderern herzlich für die bisherige Zusammenarbeit zu bedanken. Unser Dank gilt insbesondere:

- Andrew W. Mellon Foundation
- Bundesministerium für Bildung und Forschung
- Deutsche Forschungsgemeinschaft
- Europäische Kommission
- National Science Foundation
- Niedersächsisches Ministerium für Wirtschaft
- Niedersächsisches Ministerium für Wissenschaft und Kultur
- Stiftung Niedersachsen
- Union der deutschen Akademien der Wissenschaften / Akademie der Wissenschaften zu Göttingen
- VolkswagenStiftung

Göttingen, im Dezember 2013

Heike Neuroth

Inhaltsverzeichnis

Danksagung	5
-------------------	----------

Einleitung

Vorwort	13
----------------	-----------

Norbert Lossau

Die Zeit war reif	15
--------------------------	-----------

Der Aufbau der Abteilung Forschung & Entwicklung
an der SUB Göttingen

Elmar Mittler

Einführung

Open Science und Networked Science	31
---	-----------

Offenheit und Vernetzung als Leitmotive und Visionen
einer digitalen Wissenschaft im 21. Jahrhundert

Matthias Schulze und Ralf Stockmann

Bericht über das Symposium zum zehnjährigen Bestehen der Abteilung Forschung & Entwicklung	39
---	-----------

Tobias Blanke und Axel Horstmann

Forschungsprojekte

Programming for the Future?	71
------------------------------------	-----------

The Specification of User Requirements in the Design
of Virtual Research Environments for the Arts and Humanities

Kathleen M. Smith & Fotis Jannidis

Theodor Fontanes Notizbücher	85
-------------------------------------	-----------

Genetisch-kritische und kommentierte Hybrid-Edition,
erstellt mit der Virtuellen Forschungsumgebung TextGrid

Gabriele Radecke, Mathias Göbel und Sibylle Söring

**Die Edition „Johann Friedrich Blumenbach – online“
der Akademie der Wissenschaften zu Göttingen** 107

Martina Kerzel, Mike Reich und Heiko Weber

**Geisteswissenschaftliche Aufbereitung
von Daten mit TEI** 137

Werner Wegstein und Wolfgang Pempe

**From Fragments to an Integrated
European Holocaust Research Infrastructure** 157

*Tobias Blanke, Veerle Vanden Daelen, Michal Frankl,
Conny Kristel, Kepa J. Rodriguez & Reto Speck*

**Erfahrungen aus dem Projekt
„WissGrid – Grid für die Wissenschaft“** 179

Harry Enke und Bernadette Fritzschn

Lehre und Qualifizierung

Qualifizierung im Bereich digitaler Langzeitarchivierung 197

Achim Oßwald und Stefan Strathmann

Digital Humanities lehren und lernen 209

Modelle, Strategien, Erwartungen

Malte Rehbein und Patrick Sahle

Hochschule im digitalen Zeitalter 229

Die Stärkung von Informationskompetenz als Ziel
einer Empfehlung der Hochschulrektorenkonferenz

Ulrich Meyer-Doeringhaus

Technologien und Dienste

Forschungsinfrastrukturen in den Geisteswissenschaften 243

DARIAH-EU als europäische Forschungsinfrastruktur

Tobias Blanke und Christiane Fritze

Forschungsdaten-Management 257

Jens Klump und Jens Ludwig

Digitale Werkzeuge in den digitalen Geisteswissenschaften	277
Die Virtuelle Forschungsumgebung TextGrid – Status quo und neue Entwicklungen <i>Stefan E. Funk, Ubbo Veenster und Thorsten Vitt</i>	
Technische Infrastrukturen, Dienste und Services für die digitalen Geisteswissenschaften	301
<i>Peter Wittenburg und Peter Gietz</i>	
Ausblick	
Die wissenschaftliche Bibliothek im Kontext von Forschungsinfrastrukturen	325
<i>Heike Neuroth</i>	
Aus Sicht der Geisteswissenschaften: Die wissenschaftliche Bibliothek als Schnittstelle zwischen digitaler Infrastruktur und geisteswissenschaftlicher Forschung	345
<i>Andrea Rapp</i>	
Aus Sicht der Naturwissenschaften: Grid-Computing in der Teilchenphysik, Fortschritt durch Technik	355
<i>Arnulf Quadt</i>	
Verzeichnis der Autorinnen und Autoren	361
Abkürzungsverzeichnis	363

Einleitung

Vorwort

von Norbert Lossau

Informationsinfrastrukturen stellen eine wesentliche Voraussetzung für zukunftsfähige Forschung dar. Sie umfassen, analog zu Großgeräten in Forschungsinfrastrukturen, Bibliotheken, Archive, Datensammlungen, Datenbanken, objektbezogene und weitere Sammlungen. Ihre konstitutive Stellung im Wissenschaftssystem ist über die letzten zehn Jahre zunehmend in den Blickpunkt der wissenschaftspolitischen Diskussion geraten. Dem Bericht der Bund-Länder-Kommission für Bildungsplanung und Forschungsförderung zur „Neuausrichtung der öffentlich geförderten Informationseinrichtungen“ im Jahre 2006 folgten in 2011 die Empfehlungen der Kommission zur Zukunft der Informationsinfrastrukturen in Deutschland für ein „Gesamtkonzept für die Informationsinfrastruktur in Deutschland“ im Auftrag der Gemeinsamen Wissenschaftskonferenz für Bund und Länder. Wissenschaftsrat und Deutsche Forschungsgemeinschaft haben im letzten Jahr einschlägige Positionspapiere veröffentlicht.

Entscheidend für den Erfolg von Informationsinfrastrukturen in der Forschung ist die Bereitschaft und Fähigkeit zur engen Partnerschaft. Hier hat die Niedersächsische Staats- und Universitätsbibliothek mit der frühen Weichenstellung zur Gründung der Abteilung Forschung & Entwicklung im Jahre 2002 eine Modell- und Führungsfunktion im Bibliotheksbereich übernommen, die heute national und international ausstrahlt. Die Beiträge der Wissenschaftlerinnen und Wissenschaftler dieses Bandes zeugen von der engen Verbundenheit mit der Abteilung Forschung & Entwicklung und der Anerkennung für ihre herausragenden inhaltlichen und strukturbildenden Leistungen, insbesondere auch beim Aufbau von Informationsinfrastrukturen für die Geisteswissenschaften in Deutschland und Europa.

Organisationsstrukturen können nur dann erfolgreich sein, wenn sie durch Persönlichkeiten mit Leben erfüllt werden. Hier ist stellvertretend für viele Akteure Frau Dr. Heike Neuroth zu nennen, Abteilungsleiterin der ersten Stunde, und bis heute unermüdlich und mit unverminderter Kreativität aktiv. Die Niedersächsische Staats- und Universitätsbibliothek und die Universität Göttingen sind der Abteilung und ihrer Leitung zu großem Dank verpflichtet.

In einer Zeit, da die Universität die Themen Forschungs-, Informations- und IT-Infrastrukturen als integrale Aktionsfelder in ihre Zukunftsstrategie aufgenommen hat, ist die Tätigkeit der Abteilung Forschung- und Entwicklung der Bibliothek, gemeinsam mit den Forschungsabteilungen der anderen Infrastruktureinrichtungen – insbesondere der Gesellschaft für Wissenschaftliche Datenverarbeitung, GWDG – wichtiger und zukunftsweisender denn je. Den Lesern dieses beeindruckenden Zeugnisses einer langjährigen und erfolgreichen Partnerschaft von Bibliothek und Fachwissenschaft ist zu wünschen, dass sie neue Impulse auch für zukünftige Innovationen gewinnen.

Prof. Dr. Norbert Lossau

Vizepräsident der Georg-August-Universität Göttingen

Ehem. Direktor der Niedersächsischen Staats- und Universitätsbibliothek

A handwritten signature in cursive script, appearing to read "Norbert Lossau". The ink is dark and the handwriting is fluid and somewhat stylized.

Die Zeit war reif

Der Aufbau der Abteilung Forschung & Entwicklung an der SUB Göttingen

von *Elmar Mittler*

Die Forschungsbibliothek in Göttingen – Tradition mit Zukunft

Als ein perfektes Werkzeug wissenschaftlicher Forschung hat Fabian (1977)¹ die Göttinger Universitätsbibliothek des 18. Jahrhunderts bezeichnet: Der Systematische Bandkatalog spiegelt die klassifikatorische Aufstellung der Erwerbungen aus aller Herren Länder und erlaubt über das alphabetische Register den schnellen Zugriff über Verfasser oder Sachtitel. Dass Heynes Erwerbungs politik mit dem Ziel, auch aus der älteren Literatur alles zu erwerben, was den Fortschritt der Wissenschaft zeigt, vollen Erfolg hatte, stellte auch Goethe 1801 (vgl. Goethe 1892) bewundernd fest, als er nicht nur die über 50 Titel einer Liste mit Werken zur Geschichte der Farbenlehre fast vollständig erhielt, sondern man ihm weitere zur Sache dienliche Werke vorlegte, die er trotz langer Studien noch nicht kannte. Die Bibliothek bot damit mehr als die normale Rolle der Bereitstellung von nachgefragter Literatur: Dadurch, dass sie nicht nur „content“, sondern auch „context“ anbot, konnte sie eine aktive Rolle bei der Unterstützung der Forschung übernehmen. Doch dies war nicht auf die Nutzung vorhandener Literatur beschränkt. Die Universitätsbibliothek Göttingen war auch in den wissenschaftlichen Diskurs ihrer Zeit voll integriert: Ihre Neuerwerbungen wurden in den *Göttingischen Gelehrten Anzeigen* besprochen, was diesen eine weltweite Abdeckung relevanter Literatur und damit ein weit umfassenderes Rezensionsprogramm als vergleichbaren Rezensionsorganen ermöglichte. Die Idee der Forschungsbibliothek in moderne Verhältnisse zu übertragen, ist eine Herausforderung für jeden Göttinger Bibliothekar. Wichtige Aspekte der Entwicklung von 1990

¹ Vgl. auch Mittler (2004).

bis 2006 sind im Themenband der Zeitschrift *Bibliothek und Wissenschaft* „Forschungsbibliothek im Aufbruch“ dargestellt worden.²

Der Neubau der Zentralbibliothek am Platz der Göttinger Sieben 1 (offizielle Eröffnung 1993; vgl. SUB 1993) und die Einführung des internationalen Picasystems im Schulerschluss mit der Deutschen (National-) Bibliothek, der 1991 besiegelt wurde (vgl. Mittler 2001b: 30 f.), waren entscheidende Weichenstellungen für das Konzept einer offenen Bibliothek (damals sprach man gern von der „gläsernen Bibliothek“, weil durch den EDV-Einsatz die Sicht der Benutzer auch bis ins geschlossene Magazin möglich wurde) (vgl. Mittler 2008: 232 f.). Für die Studierenden waren zunächst die großen Lesebereiche und Freihandbestände, die durch Öffnung auch großer Magazinbereiche ergänzt wurden, besonders wichtig. Für den Wissenschaftler galt es, die Idee der Virtuellen Forschungsbibliothek voranzutreiben, die in digitalen wie realen Angeboten zur idealen Forschungsumgebung des Wissenschaftlers werden sollte. Dafür wurden in den im Rahmen der Sanierung des Historischen Gebäudes zu Lesebereichen umgestalteten Bereichen des Magazinbaus aus dem 19. Jahrhundert zunächst ca. 120.000 Bände Literatur in der Klassifikation des 18. Jahrhunderts frei zugänglich gemacht (vgl. Mittler 2001a; Ihlenfeld 2005);³ außerdem wurde die Göttinger Digitale Bibliothek aufgebaut, die sich in die von der Deutschen Forschungsgemeinschaft (DFG) unterstützte Entwicklung der Virtuellen Forschungsbibliothek eingliederte (vgl. Mittler 1997). Ein Entwicklungsschwerpunkt war dabei auch die gezielte Bereitstellung retrodigitalisierter Bestände im WWW: Mit der weltweit ersten Bereitstellung der Gutenbergbibel im Internet⁴ gelang dem mit Unterstützung der DFG aufgebauten Göttinger Digitalisierungszentrum ein weithin sichtbarer Leistungsbeweis (vgl. Enders/Liebetruh/Rapp 2005). Weitere Kernbestände der Bibliothek wie die große Sammlung der Reiseberichte wurden in enger Zusammenarbeit oder auf Anregung von Wissenschaftlern meist mit Unterstützung der DFG digitalisiert;⁵ bald ergaben

2 „Forschungsbibliothek im Aufbruch – Göttingen und die Bibliotheksentwicklung in Deutschland, Europa und den Vereinigten Staaten im 18. und 19. Jahrhundert“, *Bibliothek und Wissenschaft* 41 (2008), darin: Mittler (2008); vgl. auch Fast/Möller (2003).

3 Vgl. auch Glitsch/Rohlfing (2006).

4 <http://www.gutenbergdigital.de>

5 <http://www.sub.uni-goettingen.de/kopieren-digitalisieren/goettinger-digitalisierungszentrum/>

sich auch internationale Kooperationen wie bei den Sibirica mit der Library of Congress und der Russischen Nationalbibliothek Sankt Petersburg im Projekt „Meeting of Frontiers“⁶. Am intensivsten waren zunächst die gemeinsamen Aktivitäten mit den Mathematikern, die schon früh auf digitale Bereitstellung – in Deutschland zunächst im Kampf gegen den Medienbruch⁷ – drangen. In Göttingen, das aufgrund seiner langen mathematischen Tradition immer einen besonders engen Kontakt auch z.B. zur Deutschen Mathematischen Vereinigung (DMV) hatte (vgl. Becker/Habermann 2005: 214 f.), wurden auch im Rahmen der Weiterentwicklung des Sondersammelgebietes Reine Mathematik neue Wege zur Bereitstellung der eigenen wie der Bestände anderer Fachbibliotheken (nicht zuletzt der TIB Hannover) gegangen, die z.B. zur Virtuellen Fachbibliothek Mathematik⁸ führten; Digitalisierungs-, Erschließungs- und Archivierungsaktivitäten auch im internationalen Rahmen (z.B. RusDML [vgl. Schulze 2005], EMANI und EULER [vgl. Becker/Habermann (2005): 206–208]) waren Pilotprojekte, mit denen die Bibliothek ihre Rolle als Kompetenzzentrum auf dem Gebiet der Mathematik mit dem erklärten Ziel ausbaute, den Mathematikern ein Instrumentarium zu bieten, „das ihnen an ihrem Arbeitsplatz am Bildschirm möglichst vollständig die wichtigsten Hilfsmittel für die Arbeit kontinuierlich zur Verfügung stellt“ (ebd.: 215); diese Projekte strahlten auch auf andere Aktivitäten der Bibliothek aus.

2 Chancen und Probleme: Drittmittelinwerbung – Personalrekrutierung – Nachhaltigkeit

Die Einwerbung von Drittmitteln, in der sich vor allem Hans-Jürgen Becker als Meister erwies, bot Chancen wie Probleme. Sie trugen einerseits in Erweiterung der umfangreichen Drittmittel der Bibliothek, die insbesondere im Rahmen der Sondersammelgebiete über die DFG eingeworben werden konn-

6 Vgl. <http://frontiers.loc.gov/intldl/mtfhtml/mfhome.html>; http://rzblx10.uni-regensburg.de/dbinfo/detail.php?titel_id=11537&bib_id=mhh.

7 Vgl. hier vor allem die vielfältigen Publikationen von Grötschel und Lügger, die Vorkämpfer waren, z.B. Grötschel/Lügger (1996; 2007). In den USA war der Mathematiker William G. Bowens die treibende Kraft bei der Entwicklung von JSTOR (<http://www.jstor.org/>).

8 <http://vifamath.de/>; zur Gesamtentwicklung vgl. Becker/Habermann (2005).

ten, zu einer nennenswerten Steigerung der Drittmittelbilanz der Universität bei. Die SUB erreichte bald nach der Fakultät Chemie den zweiten Platz bei der Einwerbung von Fördermitteln. Andererseits führten die Projekte zu erheblichen Mehrbelastungen des Verwaltungspersonals der Bibliothek: Ohne die überragende Organisationsfähigkeit und Einsatzbereitschaft von Frau Conrad als Leiterin der Rechnungsstelle – dies sei an dieser Stelle einmal ausdrücklich erwähnt – wäre diese Entwicklung niemals zu bewältigen gewesen. Doch auch die Beantragung selbst war schon ein Problem. An DFG-Anträge war man einigermaßen gewöhnt; schon beim Bundesministerium für Forschung und Technologie (BMFT) sahen die Antragsmodalitäten ganz anders aus; die Komplexität von Anträgen bei der Europäischen Union ließ uns drei Jahre lang vergebens Schreibübungen machen, bis mit dem Pilotprojekt DIEPER⁹, mit dem Digizeitschriften¹⁰ auf europäische Basis gebracht werden sollte, ein erster Durchbruch gelang. Werner Schwartz erwies sich als ein Meister beim Umschiffen aller Klippen der europäischen Bürokratie. Das Ziel einer europäischen Allianz zur Ergänzung der amerikanischen Not-for-profit-Organisation JSTOR konnte allerdings trotzdem nicht erreicht werden. Während JSTOR bis zum Erreichen einer kritischen Masse von Digitalisaten durch die Andrew W. Mellon Foundation gefördert wurde, begnügte sich die EU einmal mehr mit der Entwicklung eines Prototypen (eine Politik, die sich sicher auch als Frucht unserer Erfahrungen inzwischen erfreulicherweise deutlich geändert hat).

Drittmittelprojekte haben in der Regel den Nachteil zeitlicher Befristung. Nur in Zeiten eines großen Angebots auf dem Arbeitsmarkt ist es relativ leicht, auch für Spezialaufgaben geeignete Bewerber zu finden. Das war bei den informationstechnisch, aber oft auch geisteswissenschaftlich anspruchsvollen Projekten nur selten der Fall. In vielen Fällen musste man begabten Newcomern die Chance zu Entwicklung zum Spezialisten geben. Das ist erfreulich oft gelungen¹¹ – allerdings bestand immer auch die Gefahr, dass die

9 <http://www.sub.uni-goettingen.de/projekte-forschung/projektetails/projekt/dieper/>

10 <https://www.digizeitschriften.de/startseite/>

11 Das herausragende Beispiel meines Nachfolgers Norbert Lossau darf an dieser Stelle einmal erwähnt werden, der als Referent für das Sondersammelgebiet Finnougristik ausgebildet wurde; nach Beendigung der Kölner Ausbildung ergab sich unerwartet eine Übergangszeit, in der er für ein von mir erworbenes DFG-Projekt zur retrospektiven Digitalisierung gewonnen werden konnte, was ihn in kurzer Zeit zum international anerkannten Spezialisten und Leiter des Göttinger Digitalisierungszentrum

erfolgreich eingearbeiteten Personen andere Posten übernahmen, bevor das Projekt zu Ende geführt worden war. Immer wieder aber war es möglich, die besten Persönlichkeiten für Dauerstellen zu gewinnen. Denn das Erfolgsrezept der Bibliothek war lange Jahre, durch Drittmittel innovative Dienstleistungen vor Ort, national und international zu entwickeln, gleichzeitig aber die traditionellen Dienste so zu rationalisieren, dass durch Ausscheiden von Mitarbeitern frei werdende Stellen immer wieder auch zur dauerhaften Erweiterung des Servicespektrums insbesondere der Digitalen Bibliothek verwendet werden konnten. Die Auszeichnung mit dem Preis „Bibliothek des Jahres 2002“ würdigte die so erreichte erfolgreiche Verbindung von Tradition und zukunftsweisenden neuen Dienstleistungen (vgl. Schmidmaier/Dämpfert 2002; Fast/Möller 2003).

Nachhaltig aber sollten auch die Forschungsaktivitäten selbst werden: Bei der engen Verzahnung der Projekte, die sich oft ergänzten und aufeinander aufbauten, war es sinnvoll, gut eingearbeitetes Personal in Folgeprojekten weiter zu beschäftigen. Das ermöglichte Kontinuität, schuf aber auch günstige Bedingungen, neue Kräfte durch einen Stamm erfahrener Mitarbeiter schnell einzuarbeiten und in das Team zu integrieren. Entscheidende Voraussetzung dafür aber war, dass Projekte erfolgreich durchgeführt und neue Ideen so projektreif gemacht werden konnten, dass sie zu weiteren Anträgen führten, die es dann bei den unterschiedlichen Förderern durchzusetzen galt. Kurz: Professionelles Management war angesagt, um diese komplexe Situation zu bewältigen. Die Abteilung Forschung & Entwicklung mit bald über 20 Mitarbeitern, die mit Heike Neuroth an der Spitze 2002 eingerichtet wurde, bestand aber zunächst fast nur aus ihr als fest angestellter Kraft – eine Herausforderung, die sie seither mit bewundernswertem Elan und hoher Frustrationstoleranz erfolgreich gemeistert und schrittweise überwunden hat.

3 Forschung in der Bibliothek

Seit den Tagen der ersten Göttinger Bibliotheksdirektoren Gesner und Heyne waren viele Bibliothekare aktive Fachwissenschaftler. Bibliotheksforschung

werden ließ. Vgl. Mittler (1998). Als weiterer führender Kopf aus dieser Zeit sei neben Heike Neuroth auch Frank Klaproth genannt. Beide haben zunächst bei geowissenschaftlichen Projekten mitgearbeitet. Durch Norbert Lossau ist auch der internationale Transfer von Göttinger Personal (insbesondere nach England) und umgekehrt gefördert worden.

hat darüber an der Göttinger Universität eine besondere Tradition. Der 1886 eingerichtete erste deutsche Lehrstuhl für Bibliothekswissenschaft wurde durch den Direktor der Bibliothek, den Altertumswissenschaftler Karl Dziatzko, besetzt. Er diente vor allem der in Preußen neu gestalteten Bibliothekarsausbildung (vgl. Pflug 2005), führte aber z.B. auch zur Herausgabe der Sammlung bibliothekswissenschaftlicher Arbeiten, in denen neben buch- und bibliothekshistorischen Arbeiten auch auf die Gegenwart bezogene Übersichten über das deutsche Bibliothekswesen und das epochemachende Etatmodell von Roquette erschienen (vgl. Roquette 1896). Die praxisorientierte Anwendung sozial- und betriebswirtschaftlicher Methoden auf die Untersuchung von Bibliotheken hatte erst in den 60er- und 70er-Jahren des vergangenen Jahrhunderts Konjunktur (vgl. Kissel 1971). Die Anwendung planungswissenschaftlicher Methoden wurde vor allem beim Bibliotheksplan Baden-Württemberg erprobt und trug mit dazu bei, dass die Bibliotheken die Chancen der allgemeinen Bildungsexpansion für ihre Entwicklung nutzen konnten (vgl. Mittler 1973/75; 1976).¹² In Ergänzung informationswissenschaftlicher Ansätze (vgl. vor allem Krieg 1973) veröffentlichte der Unterausschuss Bibliotheksforschung des Bibliotheksausschusses der DFG „Neue Perspektiven der Bibliotheksforschung“, die vor allem die praxisorientierte Forschungsarbeiten anregen wollten, wie sie seither in der 1977 neu gegründeten Zeitschrift *BIBLIOTHEK Forschung und Praxis* vorrangig publiziert wurden.¹³ Ein wichtiges Mittel zur Serviceverbesserung waren dabei immer auch Nutzerbefragungen, mit denen die in den siebziger Jahren begonnene Ausrichtung der Bibliotheken von der Sammel- zur Serviceeinrichtung systematisch unterstützt werden konnte (vgl. Stoltzenburg 1974).¹⁴ Derartige forschungsorientierte Ansätze ermöglichten bei der Umorganisation der Universität Göttingen zur Stiftungsuniversität die SUB mit rationaler Argumentation „in den universitären Steuerungskreislauf aus Zielvereinbarung, Budgetallokation und Ergebnisevaluation“ (Ceynowa 2005: 81) einzubinden.

Bibliotheksforschung und -entwicklung bekamen einen weiteren Impuls mit der wachsenden Aufgabe der Integration der elektronischen Datenver-

12 Nicht unerwähnt bleiben soll in diesem Zusammenhang: *Bibliotheksplan 1973: Entwurf eines umfassenden Bibliotheksnetzes für die Bundesrepublik Deutschland*. Berlin: Dt. Bibliothekskonferenz, 1973.

13 Neuere Perspektiven der Bibliotheksforschung. Eine Diskussionsgrundlage. In: *BIBLIOTHEK. Forschung und Praxis* 1 (1977): 3–31.

14 Vgl. aus jüngerer Zeit für Göttingen: Ceynowa et al. (2004).

arbeitung in Geschäftsgänge und Services. Die zunächst meist in Eigenentwicklung erstellten lokalen Systeme und Verbundsysteme erwiesen sich seit den 80er-Jahren zunehmend als obsolet (vgl. Becker/Müller-Dreier 2005) – der Trend, internationale Systeme einzusetzen, wurde zuerst von der Deutschen (National-) Bibliothek gemeinsam mit dem Niedersächsischen Verbund (hier unter Federführung der SUB Göttingen; vgl. Mittler 2001b) umgesetzt, wobei in Niedersachsen damit die einheitliche Einführung lokaler Systeme verbunden war (vgl. auch Mittler 2008: 233). Er machte die an den Bibliotheken aufgebauten Serviceabteilungen für die Datenverarbeitung nicht überflüssig, die aber nur in Ausnahmefällen wie in Bielefeld oder Regensburg (aber auch in der SUB Göttingen bis zur Ausgliederung des BRZN als Verbundzentrale des GBV; vgl. Becker/Müller-Dreier 2005) den Charakter von eigenständigen Forschungs- und Entwicklungsabteilungen erreicht hatten. In Göttingen ging man einen Schritt weiter, wie schon im Rahmen der Festschrift zur Verleihung des Preises „Bibliothek des Jahres 2002“ festgehalten wurde: Mit Digitalisierung, Erschließung und Bereitstellung von digitalen und gedruckten Ressourcen, elektronischem Publizieren, Interoperabilität von Metadaten und Formaten, Weiterentwicklung internationaler Standards sowie Langzeitarchivierung und Bereitstellung wurde ein breites und über lokale Bedürfnisse weit hinausgehendes Spektrum an Forschungsbereichen abgedeckt, das darauf zielte, Dienstleistungen zu entwickeln, „die allen Benutzern der SUB zur Verfügung stehen und von weiteren Einrichtungen mit- oder nachgenutzt werden“ (Fast/Möller 2003: 61) konnten.

Wie sehr die neuen Entwicklungen des Informationsmanagements der SUB auch in der strategischen Planung der Gesamtuniversität eingebracht werden konnten, zeigte 2004 der Hauptantrag GÖ* für die DFG-Förderinitiative „Leistungszentren für Forschungsinformation“, zu dem die Universität Göttingen aufgefordert worden war. Dort werden die Informationsflüsse im Informations- und Publikationssystem und die Stellung der Bibliothek in diesem Umfeld analysiert und Synergieeffekte durch Zusammenarbeit dargestellt (vgl. Koke 2004: 33–41, insbes. S. 39 f.).

4 Von der Bibliotheksforschung zur Integration der Bibliothek in die Forschung

Blieben die bisher genannten Ansätze service- oder jedenfalls bibliotheksorientiert, so gelang es der SUB auch im Rahmen der 2003 geschaffenen

Professur für Buch und Bibliothekswissenschaften in dem interdisziplinären Forschungsprojekt „Mediaconomy“¹⁵, das Forschungsfeld „Internetökonomie der Wissenschaftskommunikation“ zu übernehmen. Auch hier aber wurden konkrete Entwicklungen integriert: Als Case Study für neue Publikationsmodelle im Internet konnte der Aufbau des Universitätsverlags Göttingen¹⁶ begonnen werden.

Ziel des Aufbaus der Virtuellen Fachbibliotheken war zunächst einmal die homogene Bündelung der für ein Fach relevanten gedruckten oder digitalen Quellen in einem Portal. Durch das breite Spektrum der Sondersammelgebiete konnten hier einerseits gemeinsame Grundaspekte übertragen, aber auch differenzierte Lösungen vorangebracht werden, die sich aus den Spezialbedürfnissen der einzelnen Forschungsgebiete ergaben (vgl. Becker/Habermann 2005; Enderle 2005; Eck 2005; Schüler 2005; Pfurr 2005). So entstanden Module wie der Fach-OPAC, Online Contents, Fachinformationsführer der Online-Ressourcen (vgl. Farrenkopf 2005). Doch bald wurden auch weitere Aspekte wie der Aufbau von Dokumentenservern mit dem Ziel der Entwicklung „integrierter Servicezentren zur Unterstützung der (digitalen) Wissensproduktion (z.B. elektronisches Publizieren) und Informationsgewinnung bzw. -vermittlung (z.B. Universitätsverlag, Dokumentenserver) einbezogen. Neue Forschungsschwerpunkte werden gesetzt, da digitales Publizieren und Archivieren medien- bzw. formatneutral und plattformunabhängig realisiert werden sollte“ (Blumentritt/Neuroth 2005: 26). Dafür mussten möglichst vorhandene Standards eingesetzt oder erweitert oder neue geschaffen werden (vgl. Fischer 2005). Ein Grundprinzip war, die dafür notwendigen Entwicklungen mit Partnern in möglichst breiter nationaler oder noch besser internationaler Partnerschaft voranzubringen. Das war nicht nur bei Einzelprojekten der Fall, sondern erklärt auch das intensive Engagement in kooperativen Gruppen von Digizeitschriften¹⁷ über DIEPER¹⁸ (Zeitschriften) zu Vascoda¹⁹ (Virtuelle Forschungsbibliotheken), KOPAL²⁰ und nestor²¹ (Lang-

15 <http://www.uni-goettingen.de/de/46677.html>

16 <http://www.univerlag.uni-goettingen.de/content/about.php>

17 <http://www.digizeitschriften.de/startseite/>

18 <http://www.sub.uni-goettingen.de/projekte-forschung/projektetails/projekt/dieper/>

19 <http://de.wikipedia.org/wiki/Vascoda>

20 <http://kopal.langzeitarchivierung.de/index.php.de>; vgl. auch Neumair (2005).

21 http://www.langzeitarchivierung.de/Subsites/nestor/DE/Home/home_node.html

zeitarchivierung) bis hin zur Gründung von DINI e. V.²², dessen Geschäftsstelle (wie bei Digizeitschriften) in Göttingen angesiedelt wurde. Damit konnten außerdem wichtige Erfahrungen für das Management komplexer und internationaler Forschungs- und Entwicklungsvorhaben mit vielen Partnern gesammelt werden.

Im Falle von DINI, der Deutschen Initiative für Netzwerkinformation, in der die Serviceeinrichtungen Bibliotheken, Medien- und Rechenzentren zusammenarbeiten, wurde besonderer Wert darauf gelegt, auch die Vertreter der wissenschaftlichen Fachgesellschaften zu integrieren. Wie auch die Erfahrungen in einzelnen Göttinger Forschungsprojekten zeigten, waren die Wissenschaftler am Aufbau von Forschungsumgebungen in Zusammenarbeit mit der Bibliothek interessiert. Das galt nicht nur für die Integration von fachlichen Taxonomien mit dafür relevanten Digitalisaten (EZOLO)²³, sondern auch für den Aufbau von ASTROCAT, mit dem eine „konzeptionell neuartige Wissensmanagementlösung in Form eines netzgestützten astronomischen Katalogs erarbeitet (wurde), mit der die Datenmeldung, Erfassung und Recherche verknüpft werden (konnten)“²⁴. Neben der Verknüpfung von Daten und Literatur wurde auch in einem dreistufigen Benutzerkonzept gewährleistet, dass im Bereich Kataklysmische Veränderliche auch die Beobachtungen von weltweit verteilten Laien in die Datensammlung integriert werden konnten – eine frühe Anwendung von Crowdsourcing.

Die Möglichkeiten Virtueller Forschungsumgebungen auch für Geisteswissenschaftler wurde erstmals auf dem Wolfenbütteler Symposium „Forschung und Bibliothek“ 1996 thematisiert, bei der die Verteilte Digitale Forschungsbibliothek als „neues Paradigma von Bibliothek und Forschung“ vorgestellt wurde (vgl. Mittler 1977: 147). Zwar stand für die anwesenden Wissenschaftler zum damaligen Zeitpunkt noch der Wunsch nach der Verbesserung der Zugänglichkeit der historischen Bestände im Vordergrund, dem in Göttingen einige Jahre durch die Forschungsbibliothek im Historischen Gebäude mit freiem Zugang auch zu einem Teil der Altbestände Folge geleistet werden konnte (vgl. Glitsch/Rohlfing 2006: 32–45, insbes. S. 40); doch auch die Umsetzung der engen Verbindung von Forschung und Bibliothek in einer Virtuellen Forschungsumgebung konnte 2006 erstmals in Angriff genommen

22 <http://www.dini.de/>

23 <http://www.sub.uni-goettingen.de/projekte-forschung/projektetails/projekt/ezoolo-1/>

24 <http://www.sub.uni-goettingen.de/projekte-forschung/projektetails/projekt/astrocat-i/>

werden, wie Heike Neuroth festhielt: „In Textgrid entsteht unter der Federführung der SUB Göttingen die modulare Plattform für verteilte und kooperative Textdatenverarbeitung – der Kern eines Community Grid für die Geisteswissenschaften“ (ebd.: 62).

5 Die Zeit war reif

Versucht man abschließend ein kurzes Fazit zu ziehen, dann ist es vielleicht nicht zu viel gesagt, dass mit dem Schwerpunkt Forschung und Entwicklung in der SUB Göttingen das Erreichen des Ziels, die Bibliothek nicht nur passiv die Ergebnisse wissenschaftlicher Forschung bei Anfrage zur Verfügung stellen zu lassen, sondern sie in den Prozess der digitalen Produktion von Wissen von der Datensammlung bis zur Publikation aktiv zu integrieren, in greifbare Nähe gerückt ist. Heute sind das Entwicklungen, die in aller Munde sind. Mit Recht wird gefordert, dass sich die Bibliothek neu erfinden solle. In Göttingen sind frühzeitig erste Schritte in diese Richtung gegangen worden. Dabei konnten sicherlich nicht alle Projekte erfolgreich sein; manche Entwicklungslinie musste aus unterschiedlichen Gründen wieder aufgegeben werden. Das sind in diesem Umfeld normale Erscheinungen.

Aber die Zeit war vor zehn Jahren reif, durch die Abteilungsbildung dem Bereich Forschung und Entwicklung eine offizielle organisatorische Anerkennung geben. Damit war gewährleistet, dass die Grundlagen für viele Entwicklungen im Großen wie im Kleinen gelegt, qualifizierte Personen gewonnen und eine tragfähige Struktur zur Übernahme auch übergreifender Projekte aufgebaut werden konnten, die von der Abteilung Forschung & Entwicklung mit vielen Partnern erfolgreich durchgeführt werden konnten.

Mir selbst hat die Unterstützung von Forschung in der Bibliothek und über die Bibliothek hinaus in meinem Berufsleben viel bedeutet. Ich konnte damit viele kreative Persönlichkeiten fördern – und vielleicht auch insgesamt einen kleinen Beitrag dazu leisten, die Position der Bibliothek als dauerhafter Informationseinrichtung zu stärken, deren Aufgaben und Position bei flüchtigen Medien und massiver kommerzieller Konkurrenz immer bedeutender, aber auch gefährdeter geworden sind. Am Standort Göttingen schienen mir in den neunziger Jahren des vergangenen Jahrhunderts die besten Möglichkeiten zu bestehen, dafür mit der Aussicht auf Erfolg zu wirken.

Literaturverzeichnis

- Bargheer, Margo; Ceynowa, Klaus (Hrsg.) (2005): *Tradition und Zukunft – die Niedersächsische Staats- und Universitätsbibliothek Göttingen; eine Leistungsbilanz zum 65. Geburtstag von Elmar Mittler*. Göttingen: Univ.-Verl. Göttingen. Online: <http://webdoc.sub.gwdg.de/univerlag/2006/fsmittler.pdf>.
- Becker, Hans-Jürgen; Habermann, Katharina (2005): Die Mathematik-Projekte an der SUB. Bilanz und Perspektive. In: Bargheer/Ceynowa, S. 203–215.
- Becker, Hans-Jürgen; Müller-Dreier, Armin (2005): Von der automatisierten Zeitschriften- und Serienkatalogisierung zur Digitalen Bibliothek. Vierzig Jahre EDV-Einsatz in Erwerbung und Katalogisierung. In: Bargheer/Ceynowa, S. 117 bis 141.
- Blumentritt, Ricarda; Neuroth, Heike (2005): Die Bibliothek als Forschungseinrichtung. Entwicklungen im Bereich der Digitalen Bibliothek. SUB, Abteilung Forschung & Entwicklung. In: Bargheer/Ceynowa, S. 25–27.
- Ceynowa, Klaus (2005): Die SUB im Reformprozess der Georg-August-Universität Göttingen. In: Bargheer/Ceynowa, S. 75–82.
- Ceynowa, Klaus; Follmer, Robert; Schiel, Stefan; Tappenbeck, Inka (2004): „Ich bin eigentlich rundum zufrieden.“ – Postalische Nutzerbefragung an der Niedersächsischen Staats- und Universitätsbibliothek Göttingen. In: *BIBLIOTHEK Forschung und Praxis* 28, S. 85–98.
- Eck, Reimer (2005): Rückblicke auf 15 Jahre der jüngeren Geschichte des Sonder-sammelgebiets anglo-amerikanischer Kulturkreis und einige internationale Beziehungen. In: Bargheer/Ceynowa, S. 239–248.
- Enderle, Wilfried (2005): Virtuelle Fachbibliotheken und nationale Fachinformationsstrategie. Zu Geschichte und Konzept der Göttinger Virtual Library of Anglo-American Culture. In: Bargheer/Ceynowa, S. 217–238.
- Enders, Markus; Liebethuth, Martin; Rapp, Andrea (2005): Richtfest im Haus der Digitalen Bibliothek. Methoden – Verfahren – Werkzeuge. In: Bargheer/Ceynowa, S. 9–24.
- Fabian, Bernhard (1977): Göttingen als Forschungsbibliothek im 18. Jahrhundert. In: Raabe, Paul (Hrsg.): *Öffentliche und Private Bibliotheken im 17. und 18. Jahrhundert* (Wolfenbütteler Forschungen; Bd. 2). Bremen: Jacobi, S. 209–239.
- Farrenkopf, Stefan (2005): Virtuelle Fachbibliotheken und Vascoda. In: Bargheer/Ceynowa, S. 50–53.
- Fast, Jan-Jasper; Möller, Tobias (Red.) (2003): *Zukunft mit Tradition: Niedersächsische Staats- und Universitätsbibliothek Göttingen, Bibliothek des Jahres 2002*. Göttingen: Niedersächsische Staats- und Univ.-Bibliothek.

- Fischer, Thomas (2005): Metadaten und Formate: Fragen des Internets aus bibliothekarischer Sicht. In: Bargheer/Ceynowa, S. 42–49.
- Glitsch, Silke; Rohlfing, Helmut (Red.) (2006): *Das Historische Gebäude: Keimzelle der Georgia Augusta; 1294–2006*. Georg-August-Universität Göttingen; Niedersächsische Staats- und Universitätsbibliothek Göttingen. Göttingen.
- Goethe, Johann Wolfgang von (1892): *Tag- und Jahres-Hefte als Ergänzung meiner sonstigen Bekenntnisse von 1749 bis 1806*. [Weimarer Ausgabe 1892; Abth. 1, Bd. 35]. Weimar: Böhlau.
- Grötschel, Martin; Lügger, Joachim (1996): Neue Produkte für die digitale Bibliothek: die Rolle der Wissenschaften. In: *Die unendliche Bibliothek: Digitale Information in Wissenschaft, Verlag und Bibliothek*. Wiesbaden: Harrassowitz, S. 38–67.
- Grötschel, Martin; Lügger, Joachim (2007): Die Wissenschaftsgalaxis: Wie soll die zukünftige Informationsversorgung aussehen? In: Keil, Reinhard; Kerres, Michael; Schulmeister, Rolf (Hrsg.): *eUniversity – Update Bologna*. Münster, S. 335–354.
- Ihlenfeld, Burkard (2005): Sanierung des Historischen Gebäudes der Niedersächsischen Staats- und Universitätsbibliothek. In: Bargheer/Ceynowa, S. 9–24.
- Kissel, Gerhard (1971): *Betriebswirtschaftliche Probleme wissenschaftlicher Bibliotheken*. München: Verl. Dokumentation.
- Koke, Hartmut (Hrsg.) (2004): *GÖ* – integriertes Informationsmanagement im heterogenen eScience-Umfeld : GÖ*-Hauptantrag für die DFG-Förderinitiative „Leistungszentren für Forschungsinformation“* (GWDG-Bericht; 65). Göttingen: GWDG. Online: <http://www.gwdg.de/fileadmin/inhaltsbilder/Pdf/Publicationen/GWDG-Berichte/gwdg-bericht-65.pdf>.
- Krieg, Werner (Hrsg.): *Bibliothekswissenschaft : Versuch einer Begriffsbestimmung in Referaten und Diskussionen bei dem Kölner Kolloquium (27.–29. Oktober 1969)*. Köln: Greven.
- Mittler, Elmar (Red.) (1973/75): *Gesamtplan für das wissenschaftliche Bibliothekswesen*, hrsg. v. Arbeitsgruppe Bibliotheksplan Baden-Württemberg. München. Bd. 1: 1973, Bd. 2: 1975. Pullach b. München: Verl. Dokumentation.
- Mittler, Elmar (1976): Bibliotheksplanung. In: Kehr, Wolfgang (Hrsg.): *Zur Theorie und Praxis des modernen Bibliothekswesen*. Pullach b. München: Verl. Dokumentation, Bd. 1, S. 88–154.
- Mittler, Elmar (1977): Verteilte digitale Forschungsbibliothek – ein neues Paradigma für das Verhältnis von Bibliothek und Forschung. In: „Forschung und Bibliothek. Beiträge des Symposiums in der Herzog August Bibliothek Wolfenbüttel vom 10. bis 12. September 1976“, *Bibliothek und Wissenschaft* 30 (1977): 141–149.

- Mittler, Elmar (1997): Verteilte digitale Forschungsbibliothek – ein neuer Förderbereich der Deutschen Forschungsgemeinschaft. In: *Von Gutenberg zum Internet – 7. Deutscher Bibliothekskongress; 87. Deutscher Bibliothekartag in Dortmund 1997*. Frankfurt/Main: Klostermann, S. 81–88.
- Mittler, Elmar (Hrsg.) (1998): *Retrospektive Digitalisierung von Bibliotheksbeständen: Berichte der von der Deutschen Forschungsgemeinschaft einberufenen Facharbeitsgruppen „Inhalt“ und „Technik“* (dbi-materialien; 166). Berlin: Dt. Bibliotheksinstitut. Online: http://webdoc.sub.gwdg.de/pub/sub/2007/mitt-retr_PPN516494147.pdf.
- Mittler, Elmar (2001a): Die Göttinger Forschungsbibliothek. In: Blum, Askan (Hrsg.): *Bibliothek in der Wissensgesellschaft*. München: Saur, S. 157–164.
- Mittler, Elmar (2001b): Vom Bibliotheksrechenzentrum für Niedersachsen (BRZN) zur Verbundzentrale des Gemeinsamen Bibliotheksverbundes (VZG). In: ders. (Red.): *Zehn Jahre Pica in Niedersachsen und Deutschland: Skizzen eines Erfolgs* (Göttinger Bibliotheksschriften; 16). Göttingen: Niedersächsische Staats- und Universitätsbibliothek, S. 29–46. Online: http://webdoc.sub.gwdg.de/ebook/aw/gbs/gbs_16.pdf.
- Mittler, Elmar (2004): „Die wichtigsten Schriften aller Zeiten und Völker ... – die Göttinger Bibliothek im Zentrum einer europäischen Gelehrtenelite des 18. Jahrhunderts. In: *Georgia Augusta* 3, S. 11–17.
- Mittler, Elmar (2008): Fortschritt mit Tradition. Die SUB Göttingen auf dem Weg zur Forschungsbibliothek der Zukunft. In: Knoche, Michael (Hrsg.): *Forschungsbibliothek im Aufbruch – Göttingen und die Bibliotheksentwicklung in Deutschland, Europa und den Vereinigten Staaten im 18. und 19. Jahrhundert* (Bibliothek und Wissenschaft; 41). Wiesbaden: Harrassowitz, S. 231–240.
- Neumair, Bernhard; Ullrich, Dagmar (2005): Digitale Langzeitarchivierung – nicht nur eine technische Herausforderung. In: Erland Kolding Nielsen, Klaus G. Saur, Klaus Ceynowa (Hrsg.): *Die innovative Bibliothek*. München: Saur, S. 121 bis 138.
- Pflug, Günter (2005): Die Ausbildung des höheren Bibliotheksdienstes nach dem Zweiten Weltkrieg. In: Gerhard Hacker, Torsten Seela (Hrsg.): *Bibliothek leben – Das deutsche Bibliothekswesen als Aufgabe für Wissenschaft und Politik. Festschrift für Engelbert Plassmann zum 70. Geburtstag*. Wiesbaden: Harrassowitz, S. 80–90.
- Pfurr, Norbert (2005): Über 50 Jahre Schwerpunktbibliothek für die überregionale Literaturversorgung der Geowissenschaften – vom systematischen Zettelkatalog zur Verteilten Virtuellen Fachbibliothek GEO-LEO. In: Bargheer/Ceynowa, S. 263–289.

- Roquette, Adalbert (1896): Die deutschen Universitätsbibliotheken, ihre Mittel und ihre Bedürfnisse. In: Karl Dziatzko (Hrsg.): *Beiträge zur Theorie und Praxis des Buch- und Bibliothekswesens I* (Sammlung bibliothekswissenschaftlicher Arbeiten; 6). Leipzig, S. 40–61.
- Schmidmaier, Dieter; Dämpfert, Elke (Bearb.): *Bibliothek des Jahres: der nationale Bibliothekspreis der Jahre 2000, 2001 und 2002 / ZEIT-Stiftung Ebelin und Gerd Bucerius*, hrsg. von der Jury des Preises anlässlich der dritten Preisverleihung am 24.10.2002. Berlin: Deutscher Bibliotheksverband.
- Schüler, Mechthild (2005): Die virtuelle Kartenbibliothek – eine Momentaufnahme aus der SUB Göttingen. In: Bargheer/Ceynowa, S. 249–262.
- Schulze, Matthias (2005): Retrospektive Digitalisierung von Bibliotheksbeständen im Bereich „Verteilte Digitale Forschungsbibliothek“ – Das Projekt RusDML. In: Bargheer/Ceynowa, S. 40–42.
- Stoltzenburg, Joachim (1974): Ziele und Methoden der Projekte zur Benutzerforschung in wissenschaftlichen und öffentlichen Bibliotheken. In: *Bibliothekari-sche Kooperation: Aspekte und Möglichkeiten (Zeitschrift für Bibliothekswesen und Bibliographie; Sonderheft 18)*. Frankfurt/Main, S. 111–117.
- SUB (1993): *Niedersächsische Staats- und Universitätsbibliothek Göttingen. Dokumentation des Neubaus zur Eröffnung 1993*, hrsg. v. Staatshochbauverwaltung des Landes Niedersachsen u. Georg-August-Univ. Göttingen. Göttingen: Universitätsverl. Göttingen.

Einführung

Open Science und Networked Science

Offenheit und Vernetzung als Leitmotive und Visionen einer digitalen Wissenschaft im 21. Jahrhundert

von Matthias Schulze und Ralf Stockmann

“Powerful digital technologies for data acquisition, storage and manipulation create new opportunities, but also risk widening the ‘digital divide’. Open Science envisages optimal sharing of research results and tools: publications, data, software, and educational resources. It will rely on advanced e-infrastructures that enable online research collaboration. The potential to link cognate, and to re-use initially unrelated datasets will reveal unexpected relationships and will trigger new dynamics of scientific discovery. The collective intelligence of scientific communities will be unleashed through new collaborations across institutional, disciplinary, sectoral and national boundaries. The open science environments will help restore transparency and integrity to the scientific enterprise, for all to see.” (ALLEA 2012: 5)

Dieser kurze Auszug aus der Präambel der Erklärung „Open Science for the 21st century“, die der Zusammenschluss der Europäischen Akademien (ALLEA – ALL European Academies, The European Federation of National Academies of Sciences and Humanities) im April 2012 verabschiedet hat, macht deutlich, dass die Forderung nach offenem Umgang mit Wissen und Information im wissenschaftlichen Bereich zunehmend an Relevanz gewinnt.¹

Mehr oder weniger eingeführte Begriffe, Initiativen und Schlagworte prägen zunehmend den öffentlichen Diskurs, wenn es um Wissenschaftskommunikation, wissenschaftliche Infrastruktur und wissenschaftliches Arbeiten geht: Open Source, Open Access, Open Review, Open Knowledge, Open Metrics, Open Data etc., um nur einige zu nennen. Das gemeinsame Präfix „offen“ beinhaltet dabei zwei unterschiedliche Facetten: Zum einen die Offenheit des Zugangs zu Daten, Code oder Ergebnissen, zum anderen das Ge-

¹ Daneben erschienen zur selben Zeit von international angesehenen Einrichtungen weitere Berichte, Erklärungen und Positionspapiere. Hingewiesen sei hier beispielhaft nur auf The Royal Society (2012).

bot der Transparenz, also der Offenlegung von Verfahren, Methoden und Zielen.² Neben der Offenheit von Wissenschaft ist als ein weiteres Paradigma, das sich immer stärker bemerkbar macht und zur Geltung gelangt, die zunehmende Vernetzung von online betriebener Wissenschaft und Forschung zu konstatieren. Etwa anderthalb bis zwei Jahrzehnte, nachdem das Internet bzw. das Digitale allgemein in der Wissenschaft bzw. der Wissenschaftskommunikation Einzug gehalten haben, wird die Vernetzung von Wissenschaft als ein dramatischer und grundlegender Wandel begriffen und beschrieben.³

Das Internet, Online-Kommunikation und Online-Verfahren revolutionieren nicht nur wirtschaftliche Prozesse, den (sozialen) Alltag der Menschen und weitere Lebensbereiche, sondern zunehmend auch Wissenschaft und Forschung. Einige der immer wieder in diesem Zusammenhang genannten Begrifflichkeiten sind Virtuelle Forschungsumgebungen, Forschungsdaten, Grid, Clouds usw. Ob und inwieweit diese Schlagworte und Begriffe, die dahinter stehenden Tendenzen und Initiativen am Beginn des 21. Jahrhunderts die Wissenschaft prägen und damit einen grundlegenden Strukturwandel hin zu mehr Offenheit, also zu „Open Science“, und zu Vernetzung, also zu „Networked Science“, Ausdruck und Geltung verschaffen und welche Rahmenbedingungen hierfür noch geschaffen werden müssen, das ist aus heutiger Sicht und Perspektive noch nicht wirklich absehbar, aber es gibt einige Anhaltspunkte und Indizien, die dies hoffen lassen. Im Folgenden sollen anhand der eingangs zitierten Europäische-Akademien-Erklärung einige dieser Tendenzen skizziert und benannt werden.

2 Eine sehr gute Einführung und einen ersten deutschsprachigen Überblick über Initiativen und Bewegungen, die offenen und möglichst auch einfachen Zugang zu (wissenschaftlichen) Informationen fordern oder bereits gewährleisten, liefert der von Ulrich Herb herausgegebene Sammelband *Open Initiatives: Offenheit in der digitalen Welt und Wissenschaft* (Herb 2012b), der einige lesenswerte Beiträge zum Thema Open Science und zu verwandten Themenfeldern bzw. Initiativen enthält. Besonders erwähnenswert im Zusammenhang mit der Thematik Wissenschaft und Offenheit sind die Beiträge von Daniel Mietchen (Mietchen 2012), Jens Klump (zu Forschungsdaten), Jutta Haider (zu Open Access) und Ulrich Herb (Herb 2012a) selbst.

3 Einen instruktiven und wichtigen Beitrag zu dieser Thematik liefert Michael Nielsen mit seiner kürzlich erschienenen Untersuchung, die im Titel gar von einer neuen Ära der vernetzten Wissenschaft handelt: *Reinventing Discovery. The New Era of Networked Science* (Nielsen 2012).

Um diese „Vision for Open Science“ Realität werden zu lassen, bedarf es ohne Zweifel noch einiger Anstrengungen und der Schaffung von Voraussetzungen. Diese – auch als „digitale Transformation“⁴ bezeichnete – Entwicklung, wirkt sich dabei natürlich nicht nur auf die Zugänglichkeit und Verarbeitungsform der Inhalte und der Informationsinfrastruktur aus, sondern auch auf die Wissenschaft und alle mit ihr in Zusammenhang stehenden Bereiche und Prozesse selbst (etwa im Bereich der Kommunikations-, Arbeits-, Erkenntnis- und Publikationsprozesse).⁵ Die Erklärung der Europäischen Akademien selbst nennt im nachfolgenden Text hierfür drei zu erfüllende wesentliche Bedingungen: „Open Scientific Content“, „Open e-Infrastructures“ und „Open Scientific Culture“ (vgl. ALLEA 2012: 5). Während es sich bei der letzten der drei Bedingungen eher um „sozio-kulturelle“ Arbeits-/Verhaltensweisen und Verfahren handelt, geht es bei den beiden anderen Voraussetzungen um Offenheit bezogen auf Inhalte (explizit ist in der Erklärung von Publikationen und Daten die Rede, aber auch von Lehrmaterialien und Software) sowie vor allem auch infrastrukturelle Gegebenheiten.

„Open Scientific Culture“

Wie bei allen gesellschaftlichen Transformationsprozessen entstehen bei der Umsetzung einer „Open Scientific Culture“ verschiedene Geschwindigkeiten. Auch unter Wissenschaftlern finden sich „early adopter“, die sich seit Jahren etwa auf Konferenzen via Twitter-Hashtags vernetzen. Wikis haben breiten Einzug gehalten in den Wissensmanagement-Alltag von Wissenschaftlern, selbst Facebook wird nahezu subversiv genutzt zur Organisation von kleinen Arbeitsgruppen – in Ermangelung ähnlich bedienfreundlicher und verfügbarer Forschungsumgebungen. So wie die E-Mail seit Jahren aus keinem Wissenschaftsbereich mehr wegzudenken ist, werden auch diese

4 so auch der Titel eines aktuellen DFG-Positionspapiers: „Die digitale Transformation weiter gestalten – Der Beitrag der Deutschen Forschungsgemeinschaft zu einer innovativen Informationsinfrastruktur für die Forschung“, Juli 2012: http://www.dfg.de/download/pdf/foerderung/programme/lis/positionspapier_digitale_transformation.pdf

5 Diese „digitale Transformation“ betrifft sicherlich alle wissenschaftlichen Disziplinen, allerdings wohl nicht in gleichem Maße und vergleichbarer Intensität und Ausprägung. Einen sehr lesenswerten Beitrag zur Thematik „neue Werkzeuge und vernetzte Wissenschaft“ (mit Beispielen aus geisteswissenschaftlichen Disziplinen wie der Germanistik und der Philosophie) lieferten Horstmann, Kronenberg und Neubauer (2011).

neuen, webbasierten Dienste eine immer weitere Verbreitung finden und zunehmend als selbstverständlich angesehen werden.

Für die Einführung von dezidiert für die Wissenschaft entwickelten Diensten, etwa fachspezifischen Virtuellen Forschungsumgebungen, stellen die unterschiedlichen Geschwindigkeiten bei der Technologieaneignung jedoch ein Problem dar. Selbst wenn Wissenschaftler eng in die Entwicklung eingebunden werden, so werden sich in den jeweiligen Communities doch Widerstände gegen neue Verfahren und Methoden regen. Grundsätzlich unterscheidet man nun zwei Strategien: Top-down durch Förderstrategien, Vorgaben und Empfehlungen oder Bottom-up durch Graswurzelprojekte und den Einsatz von „Evangelists“. In beiden Fällen steht und fällt der Erfolg damit, ob sich der jeweiligen Zielgruppe ein unmittelbarer Mehrwert und Nutzen erschließt.

„Open Scientific Content“

Die Entwicklungen hinsichtlich der für die Wissenschaft relevanten Informationen ist spätestens seit etwa Mitte der 1990er-Jahre davon geprägt, dass wissenschaftlich relevante Informationen nur noch oder vor allem digital vorliegen und produziert werden bzw. durch Digitalisierung bisher analoger Materialien alte Wissensbestände digital verfügbar gemacht werden. Hinzu kommt auch, dass die Menge an Informationen und Inhalten, seien sie „digital-born“ oder im Nachhinein digitalisiert, immer schneller zunimmt.

Dies basiert u.a. auch auf einem Phänomen, welches zur selben Zeit verstärkt zum Tragen kam und sich zunehmend verbreitete: die sogenannte „E-Science“, bzw. neuerdings eher als „E-Research“ bezeichnet. Wesentlich für diese Entwicklung der Wissenschaft (das „E“ kann in diesem Zusammenhang am ehesten für „enhanced“, also angereichert bzw. verbessert stehen) waren technisch/strukturelle Faktoren. Hierunter sind die zunehmende Vernetzung und Anwendung von verteilten Rechnerarchitekturen, anwachsende Rechenressourcen, Nutzung von Netzwerktechnologien an verteilten Standorten zu fassen (ein wichtiges Schlagwort in diesem Zusammenhang ist das „Grid-Computing“ und ein weiteres die „Cloud“). Wesentlich hierbei ist die Tatsache, dass Forscher zunehmend Netzwerkstrukturen nutzen wollen, ohne sich mit den Rechnerarchitekturen auseinandersetzen zu müssen.⁶

⁶ Dies ist sicherlich auch ein Grund dafür, dass sich zunehmend auch die Geisteswissenschaften (Stichwort „Digital Humanities“) in diesem Bereich engagieren.

Neben den digitalen oder digitalisierten Inhalten gewinnt – wie oben schon erwähnt – auch die infrastrukturelle Ebene immer mehr an Bedeutung. Mit Blick auf die infrastrukturellen Implikationen ist natürlich vor allem und in erster Linie das Internet selbst zu nennen, das seit spätestens Mitte der 1990er-Jahre entscheidend Einfluss auch auf die Wissenschaftskommunikation und das wissenschaftliche Arbeiten genommen hat. Die Inhalte werden nicht nur immer mehr und vor allem in digitaler Form konsumiert, sondern auch kollaborativ und kooperativ, zeitlich versetzt, durch teilweise räumlich weit verstreute Arbeitsgruppen und Forschungsverbände, genutzt und weiterverarbeitet. Die digital betriebene Wissenschaft ist wesentlich auf Arbeitsweisen und infrastrukturelle Gegebenheiten angewiesen, die Computer-unterstützt, virtuell und Internet-basiert geprägt sind (vgl. Horstmann/Kronenberg/Neubauer 2011).

„Open E-Infrastructures“

Die Diskussion über die Informationsinfrastruktur für die Wissenschaft und Forschung wird auch in Deutschland seit einiger Zeit auf verschiedenen Ebenen und in unterschiedlichen Zusammenhängen sehr intensiv geführt. Um nur drei Beispiele aus der jüngsten Zeit zu nennen: Im April 2011 hat die von der Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder (GWK) beauftragte Kommission „Zukunft der Informationsinfrastruktur“ (KII) ein „Gesamtkonzept für die Informationsinfrastruktur in Deutschland“ vorgelegt.⁷ Daneben hat die Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen ebenfalls Handlungsfelder identifiziert und Lösungen vorgeschlagen.⁸ Ergänzt wurden diese beiden Initiativen durch Empfehlungen des Wissenschaftsrats (WR), der auch im Jahre 2011 „Übergreifende Empfehlungen zu Informationsinfrastrukturen“ abgegeben hat.⁹

Ohne im Einzelnen auf die Ergebnisse dieser teilweise sehr umfangreichen Papiere und Empfehlungen einzugehen, ist doch allen Initiativen gemeinsam, dass sie – sich teilweise ergänzende – Handlungsfelder identifiziert haben, auf die zukünftig das Augenmerk aller Akteure im Wissenschaftspro-

7 http://www.allianzinitiative.de/fileadmin/user_upload/KII_Gesamtkonzept.pdf

8 <http://www.allianzinitiative.de/de/>

9 <http://www.wissenschaftsrat.de/download/archiv/10466-11.pdf>

zess (seien es die Wissenschaftler selbst, die wissenschaftlichen Einrichtungen, die Forschungsförderer oder auch Infrastruktureinrichtungen) legen müssen. Beispielhaft seien hier die acht von der KII benannten Handlungsfelder genannt: Lizenzierung, Hosting/Langzeitarchivierung, nichttextuelle Materialien, Retrodigitalisierung/kulturelles Erbe, Virtuelle Forschungsumgebungen, Open Access, Forschungsdaten, Informationskompetenz/Ausbildung.¹⁰

In unmittelbarem Zusammenhang mit diesen Handlungsfeldern benennt die Kommission (KII) zudem als prinzipielle Aufgabe für die Informationsinfrastruktur, „... Rahmenbedingungen dafür zu schaffen,

- dass digital vorliegende wissenschaftliche Information jedweden Typus (z.B. Publikationen, Datenbanken, Forschungsdaten) jederzeit und von überall aus für den Nutzer verfügbar sind
- dass die Nutzer darüber hinaus die Information möglichst integriert in ihre jeweils aktuellen Arbeitszusammenhänge und Arbeitsumgebungen einbauen, weiterbearbeiten und kollaborativ nutzen und nachnutzen können
- dass die in den neuen Arbeitsumgebungen erzielten Ergebnisse wieder in den Prozess der wissenschaftlichen Wertschöpfung zurückfließen.“¹¹

Neben der Art und der Ausprägung der Handlungsfelder ist auch dieses Aufgabenspektrum eindeutig darauf ausgerichtet, analog zur Open-Science-Erklärung der Europäischen Akademien, eine Vision zu entwickeln, die im Kern darauf abzielt, wissenschaftliche Inhalte für Nutzer frei verfügbar („jederzeit und von überall verfügbar“, „kollaborativ nutzen und nachnutzen“) bereitzustellen. Auch wenn also die Begrifflichkeiten „offen“, „Offenheit“, „open“ hier nicht explizit genannt werden, so ist die Tendenz doch eindeutig: Ähnlich wie die oben genannten internationalen Papiere setzen auch die deutschen Wissenschaftseinrichtungen (bzw. die beteiligten Förder- und Infrastruktureinrichtungen) auf das Potenzial der Offenheit und Vernetzung bzw. deren perspektivische Umsetzung.

10 http://www.allianzinitiative.de/fileadmin/user_upload/KII_Gesamtkonzept.pdf, S. 25.

Das Maß der Übereinstimmung zwischen den unterschiedlichen Empfehlungspapieren und deren ähnliche Ausrichtung zeigt sich schon daran, dass die Schwerpunkttinitiative „Digitale Information“ fünf dieser acht Handlungsfelder (nämlich Lizenzierung, Hosting, Forschungsdaten, Open Access und Virtuelle Forschungsumgebungen) ebenfalls ausgemacht hat.

11 ebd., Z. 336–343

Ausblick

Allerdings sollte auch nicht außer Acht gelassen werden, dass die nun proklamierte Offenheit und Transparenz der Wissenschaft nicht überall auf fruchtbaren Boden fällt. Selbst bei jungen Wissenschaftlerinnen und Wissenschaftlern herrschen teils diffuse, teils reale Ängste: Was wird aus meinem Impact-Faktor, wenn ich Open Access publiziere? Wird die These meiner Promotion von jemand anderem schneller bearbeitet, wenn ich sie zu früh transparent mache? Sind technologische Werkzeuge wirklich nachhaltig und sinnvoll, verkürzen sie vielleicht die immer noch benötigten und sinnvollen „tiefen Gedanken“ durch verführerische, schnelle Abkürzungen?

Solche Überlegungen lassen sich schwerlich als reine Maschinenstürmerei diskreditieren oder als stumpfe Fortschrittsfeindlichkeit abtun. Vielmehr sind sie begründeter Ausdruck einer jahrhundertealten Wissenschaftsphilosophie, in der immer auch das auf Konkurrenz ausgelegte, klandestine Handeln belohnt wurde. Zudem zeigt sich immer wieder, wie unterschiedlich die Erwartungen und Befindlichkeiten sich in den verschiedenen Wissenschaftsdisziplinen gestalten.

Wenn wir daher über die nächsten strategischen Schritte nachdenken, um Open Science als Paradigma zum Erfolg zu führen, reicht es nicht, sich nur auf Geldgeber, Förderprogramme und Antragsprosa zu konzentrieren. Gefragt sind auch Moderatoren, Evangelisten – ja, sogar Seelsorger –, die die Befindlichkeiten der verschiedenen, noch miteinander fremdelnden Partner wie Wissenschaftler, Bibliotheken, Rechenzentren, Infrastrukturdienstleister oder Softwareentwickler austarieren.

Es wäre zu wünschen, dass in diesem neuen, hochinteressanten Zusammenspiel auch eine Prise Mut zum Experiment und zum Risiko belohnt und gefördert würde.

Literaturverzeichnis

ALLEA (2012): Open Science for the 21st century. A declaration of ALL European Academies presented at a special session with Mme Neelie Kroes, Vice-President of the European Commission, and Commissioner in charge of the Digital Agenda on occasion of the ALLEA General Assembly held at Accademia Nazionale dei Lincei, Rome, on 11–12 April 2012. <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/allea-declaration-1.pdf>.

- Herb, Ulrich (2012a): Offenheit und wissenschaftliche Werke: Open Access, Open Review, Open Metrics, Open Science & Open Knowledge. In: ders. (Hrsg.): *Open Initiatives: Offenheit in der digitalen Welt und Wissenschaft*, Saarbrücken: Universaar, S. 11–44. Online: http://scidok.sulb.uni-saarland.de/volltexte/2012/4866/pdf/Herb_mit_Deckblatt.pdf.
- Herb, Ulrich (Hrsg.) (2012b): *Open Initiatives: Offenheit in der digitalen Welt und Wissenschaft*, Saarbrücken: Universaar. Online: <http://universaar.uni-saarland.de/monographien/volltexte/2012/87/>.
- Horstmann, Wolfram; Kronenberg, Hermann; Neubauer, Karl Wilhelm (2011): Vernetzte Wissenschaft. Effektivere Forschung mit neuen Werkzeugen. In: *B.I.T. online* 14, 2011, Nr. 4, S. 354–362. Online: <http://www.b-i-t-online.de/heft/2011-04/fachbeitrag-horstmann.pdf>.
- Mietchen, Daniel (2012): Wissenschaft zum Mitmachen, Wissenschaft als Prozess: Offene Wissenschaft. In: Ulrich Herb (Hrsg.): *Open Initiatives: Offenheit in der digitalen Welt und Wissenschaft*. Saarbrücken: Universaar, S. 55–64. Online: http://eprints.rclis.org/bitstream/10760/17217/1/Mietchen_mit_Deckblatt.pdf.
- Nielsen, Michael (2012): *Reinventing Discovery. The New Era of Networked Science*. Princeton/Oxford.
- The Royal Society (2012): Science as an open enterprise. http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf.

Bericht über das Symposium zum zehnjährigen Bestehen der Abteilung Forschung & Entwicklung

von Tobias Blanke und Axel Horstmann

I.



Abb. 1 Die Symposiumsteilnehmer

„Zehn Jahre Abteilung Forschung & Entwicklung der Niedersächsischen Staats- und Universitätsbibliothek Göttingen“ – bei einem Symposium-Titel dieser Art erwartet man vor allem Rückschau, Bestandsaufnahme und Bewertung mit Blick auf das Geleistete, vielleicht auch einen mehr oder minder mutigen Blick auf künftige Herausforderungen und Chancen für das eigene Wirken. Das Göttinger Symposium, über das hier zu berichten ist, bot seinen über 100 Teilnehmern weit mehr als das. Denn es hat die Arbeit der 2002 gegründeten Abteilung von vornherein in den größeren Zusammenhang der Entwicklung der Informationsinfrastruktur und insbesondere der Beziehungen zwischen Bibliothek und Fachwissenschaft gestellt – ganz so, wie es der vollständige Titel der Veranstaltung versprach. Es war die Frage nach den grundlegenden Veränderungen auf diesem Feld, die den roten Faden des

Symposiums bildete: nach den Veränderungen der akademischen Welt durch den Wandel der Informations- und Wissensgesellschaft und nach den neuen Aufgaben und Rollen für Wissenschaftliche Bibliotheken, nach Veränderungen bei den Infrastruktur- und Serviceeinrichtungen ebenso wie nach neuen Formen und Möglichkeiten wissenschaftlicher Arbeit und nicht zuletzt nach der Zukunft des Zusammenwirkens zwischen den Infrastruktureinrichtungen auf der einen und den Fachwissenschaften auf der anderen Seite. Natürlich lässt sich ein solch ‚weites Feld‘ nicht in vier Stunden vollständig aus- und durchmessen – nicht einmal in Göttingen –, doch – und darauf kam es letztlich an – es gelang, das breite Spektrum der vielschichtigen Probleme, aber auch der vielfältigen Chancen und ermutigenden Perspektiven, die dieses weite Feld zu bieten hat, ebenso anschaulich wie eindringlich vor Augen zu führen und dabei zugleich Anregungen und Impulse für die weitere Arbeit zu geben. Dafür gilt es allen Beteiligten und namentlich den Veranstaltern mit ihren vielen Helferinnen und Helfern schon an dieser Stelle Dank und Anerkennung auszusprechen. Der Weg nach Göttingen hat sich gelohnt!

II.

In ihrer Eröffnung und Einführung unterstrichen Uwe Schwiiegelshohn und Norbert Lossau die beeindruckende Leistung, die die Abteilung Forschung & Entwicklung in den zehn Jahren ihres Bestehens erbracht habe, und dankten der Abteilungsleitung, allen voran Heike Neuroth, für ihr großes Engagement und ihre höchst erfolgreiche Arbeit. Heike Neuroth sei nicht nur der Motor der Abteilung, sondern auch für die SUB Göttingen insgesamt als Person und Wissenschaftlerin richtungsweisend.

Bibliotheken – so Uwe Schwiiegelshohn – gehörten zu den ältesten Serviceeinrichtungen im akademischen Umfeld und tendierten als solche in der Regel dazu, auf Althergebrachtem zu beharren. Umso höher sei es der SUB Göttingen anzurechnen, dass sie eine eigene Abteilung für Forschung und Entwicklung eingerichtet habe, die neue Entwicklungen aufgreife, selber maßgeblich vorantreibe und dabei versuche, zwischen den verschiedenen Disziplinen und Fachbereichen zu vermitteln, Gemeinsamkeiten zu finden und auf diese Weise die Forschung bei ihrer Arbeit an neuen Themen, Fragestellungen und Problemen gezielt zu unterstützen.



Abb. 2 Prof. Dr. Norbert Lossau,
SUB Göttingen

Norbert Lossau würdigte in diesem Zusammenhang insbesondere auch die Rolle des damaligen Direktors der SUB Elmar Mittler bei der Gründung der Abteilung. Ohne seine Weitsicht und tatkräftige Unterstützung gäbe es heute kein Jubiläum zu feiern. Gegenwärtig treibe die Abteilung, die ihrerseits auf entsprechenden Vorarbeiten seit Mitte der 1990er-Jahre aufbauen konnte, nicht nur Forschung und Entwicklung im Zusammenwirken mit der Fachwissenschaft voran, sondern verfüge auch über eigene Kompetenzzentren.

Zentraler Stellenwert komme in der Arbeit der Abteilung dem Community Building zu, vor allem in und mit den Geisteswissenschaften. Das Projekt TextGrid¹ liefere hier ein prominentes Beispiel. Allerdings werde auch über die Geisteswissenschaften hinaus fächerübergreifend zusammengearbeitet, so etwa in Projekten wie WissGrid² oder WissNetGrid³. Diese Kooperationen wolle die SUB weiter fortführen. Aus diesem Grunde habe man auch die Einrichtung des Göttingen Centre for Digital Humanities (GCDH)⁴ nachdrücklich unterstützt und auch die Kooperation mit der Akademie der Wissenschaften befördert. Darüber hinaus sei die SUB bestrebt, die Zusammenarbeit auch mit anderen Disziplinen zu intensivieren und sich verstärkt im Bereich E-Research zu engagieren. Dass diese Aktivitäten inzwischen auch in der Strategieentwicklung der Universität Göttingen ‚angekommen‘ seien und dort entsprechend wahrgenommen würden, bedeute der SUB sehr viel. So habe sich die übergreifende Zusammenarbeit von Universität, Bibliothek, GCDH, Max-Planck-Instituten

1 <http://www.textgrid.de>

2 <http://www.wissgrid.de>

3 <http://www.wisnetgrid.org>

4 <http://www.gcdh.de/de>

und der Gesellschaft für Wissenschaftliche Datenverarbeitung Göttingen (GWDG)⁵ inzwischen zu einem Alleinstellungsmerkmal dieses Standorts entwickelt. In dieser Richtung werde man nach Kräften weiterarbeiten. Nur so ließen sich die bei der Weiterentwicklung der Informationsinfrastruktur anstehenden Aufgaben bewältigen. Angesichts dieser Stärken des Standorts Göttingen wäre auch über die Einrichtung eines eigenen Instituts für Informationsinfrastrukturen nachzudenken. Einstweilen sei das aber noch Zukunftsmusik.

III.

Im dem von Norbert Lossau moderierten ersten Tagungsabschnitt referierten Ulrich Meyer-Doerpinghaus von der Hochschulrektorenkonferenz (HRK) über „Die Stärkung der wissenschaftlichen Informationskompetenz als hochschulpolitische Aufgabe“ und im Anschluss daran Achim Oßwald von der Fachhochschule Köln zum Thema „Auf der Suche nach dem Alleinstellungsmerkmal: Neue (?) Aufgaben und Rollen für Wissenschaftliche Bibliotheken“.

Ulrich Meyer-Doerpinghaus zufolge steht „Informationskompetenz“ seit einiger Zeit im Fokus der Hochschulrektorenkonferenz. Schon bei der Arbeit der „Kommission Zukunft der Informationsinfrastruktur (KII)“ an der Entwicklung eines entsprechenden Gesamtkonzepts für Deutschland habe sich die Bedeutung dieses Themas gezeigt. Gegenwärtig vollzögen sich im individuellen wie im gesellschaftlichen Umgang mit Information rasante Veränderungen. So seien digitale Information und Social Media inzwischen allgegenwärtig. Das gehe einher mit einem kulturellen Wandel im Umgang mit Information, bei dem nun sensorische (Beispiel: iPod) und partizipative Elemente (Beispiel: Wikis) eine immer wichtigere Rolle spielten. Dieser Wandel mache auch vor den Hochschulen nicht halt. Dementsprechend müsse man beim Thema Informationskompetenz auch die gesamte Hochschule – d.h. neben Forschung, Lehre und Studium auch die Leitungs- und administrative Ebene – einbeziehen. Von diesen neuen Entwicklungen könne die Hochschule namentlich in der Lehre erheblich profitieren. Dies setze allerdings voraus, dass man die Studierenden, für die die Nutzung von iPhone und Twitter zur Normalität gehöre, bei ihrem gewohnten Umgang mit Informationsmedien abhole.

5 <http://www.gwdg.de/index.php>



Abb. 3 Dr. Ulrich Meyer-Doeringhaus, HRK

Für Meyer-Doeringhaus bezeichnet Informationskompetenz – so wörtlich – „die Gesamtheit aller Fähigkeiten und Fertigkeiten, die erforderlich sind, um situationsrelevante Informationsbedarfe festzustellen, Informationen zu beschaffen, weiterzuverarbeiten, zu bewerten, zu präsentieren und Nutzungsbedingungen einzuordnen. Dabei sind neue Darstellungsformate, insbesondere im Bereich der Informationsvisualisierung, eingeschlossen“. Als Teilfertigkeiten gehörten dazu technische, kommunikative, soziale, organisatorische und disziplinspezifische Kompetenz. Diesem Verständnis nach gehe Informationskompetenz über bloße Medienkompetenz wesentlich hinaus.

Was Informationskompetenz in Lehre und Studium betreffe, so werde sie in den Hochschulen inzwischen an zahlreichen Stellen gelehrt und gelernt. Dies sei auch dringend nötig. Denn Studierende müssten befähigt werden, Informationen technisch aufzubereiten, in ihre persönliche Informationsumgebung zu übernehmen, didaktisch ausgereift zu präsentieren und langfristig zu archivieren. Dabei stellten sich in der Lehre mehrere Herausforderungen: Erstens müsse das Lehrangebot bereits vorhandene Kompetenzen berücksichtigen, was wiederum eine genaue Vorstellung von dem voraussetze, was die Studierenden ihrerseits bereits mitbrächten. Zweitens gebe es noch viele Defizite in der Abstimmung der Angebote der verschiedenen Akteure; In-

formationskompetenz werde zwar vielerorts vermittelt, es fehle jedoch an Koordination. Drittens lasse die curriculare Einbindung bestehender Aktivitäten noch zu wünschen übrig. Die meisten Angebote stammten von Bibliotheken und Rechenzentren, doch seien nur wenige direkt ins Studium integriert und ECTS-fähig. Dies müsse sich ändern, damit jeder Studierende automatisch mit dem Thema konfrontiert werde. Für die HRK resultierten daraus zwei zentrale Forderungen: zum einen die curriculare Einbindung einschlägiger Angebote, zum anderen die flächendeckende Einführung entsprechender Querschnittsmodule mit generischem Anteil im Bachelor- und fachspezifischem Anteil im Master-Studium, jeweils ausgerichtet am Bedarf der Studierenden unter Einbeziehung neuer Entwicklungen. Für die Lehrenden und ihre Qualifizierung bedeute dies, dass sie mit dem, was Studierende mitbrächten, vertraut und bereit sein müssten, mit relevanten Akteuren wie Bibliotheken und Rechenzentren zu kooperieren und sich in Themen wie Web 2.0, interaktive Lehr- und Lernszenarien, Umgang mit Literaturverwaltungsprogrammen in den Geisteswissenschaften oder die Verwaltung eines elektronischen Semesterapparates soweit einzuarbeiten, dass sie die damit verbundenen Möglichkeiten in der Lehre entsprechend nutzen könnten. Dabei müssten die Lehrenden allerdings von ihren Hochschulleitungen unterstützt werden – und zwar wesentlich besser als bisher.

In der Forschung habe sich Informationskompetenz bislang nur in einigen wenigen Fachkulturen etabliert. Allerdings zeichneten sich auch hier unumkehrbare Trends ab. Stichworte seien: Virtuelle Forschungsumgebungen, elektronisches Publizieren, Open Access, Forschungsdaten, Digitalisierung, Langzeitarchivierung. Um die Forscherinnen und Forscher für die Arbeit mit diesen neuen Instrumentarien und technischen Möglichkeiten hinreichend zu qualifizieren, fordere die HRK,

- die genannten Themen stärker als bisher in die Graduierten- und Postgraduiertenausbildung einzubinden,
- entsprechende Kompetenz- und internationale Forschungsnetzwerke zu fördern und zu unterstützen,
- das Qualifikationsprofil der Mitarbeiterinnen und Mitarbeiter von Bibliotheken um neue Berufsbilder wie die des Data Curator und Data Librarian zu erweitern.

Letztere müssten fachwissenschaftlich hochkompetent und zugleich imstande sein, auch komplette Informationsstrukturen zu steuern. Hier seien neben den

Bibliotheken auch die Rechenzentren gefordert, die zu Themen wie Data Mining oder Simulation wichtige Beiträge leisten könnten (vgl. HRK 2012).

Das Stichwort „Bibliotheken“ lieferte den nahtlosen Übergang zum Vortrag von Achim Oßwald über „Neue (?) Aufgaben und Rollen für Wissenschaftliche Bibliotheken“. Oßwald verwies zunächst auf die Arbeit von „nestor“⁶, einem Zusammenschluss von mittlerweile zwölf Hochschulen aus Deutschland, Österreich und der Schweiz,⁷ die sich bei der Aus-, Fort- und Weiterbildung im Bereich digitaler Langzeitarchivierung engagieren. Diese Aktivitäten, die koordiniert und immer wieder vorangetrieben würden von der Abteilung Forschung & Entwicklung der SUB Göttingen, umfassten:

- die Entwicklung von E-Learning-Tutorials⁸ zur digitalen Langzeitarchivierung,
- das *nestor Handbuch* (Neuroth et al. 2009; 2010) sowie eine Bestandsaufnahme zur *Langzeitarchivierung von Forschungsdaten* (Neuroth et al. 2012),
- die seit 2007 regelmäßig stattfindenden *nestor Schools*⁹,
- die Mitarbeit im Projekt Digital Curator Vocational Education Europe (DigCurV)¹⁰, in dem ein curriculärer Rahmen für die Weiterbildung im Bereich der digitalen Langzeitarchivierung entwickelt werde.

Hier sei vor allem Heike Neuroth und Stefan Strathmann für ihr großes Engagement zu danken.

Welche Vorteile biete es für eine Bibliothek, über eine eigene Forschungs- und Entwicklungsabteilung zu verfügen? Oßwald zufolge sind es vier Leistungsbereiche, in denen eine solche Abteilung ihr Potenzial entfalten könne, indem sie

6 http://www.langzeitarchivierung.de/Subsites/nestor/DE/Home/home_node.html

7 Archivschule Marburg, Fachhochschule Chur, Fachhochschule Köln, Fachhochschule Potsdam, Hochschule Darmstadt, Hochschule der Medien Stuttgart, Hochschule für Technik und Wirtschaft Berlin, Hochschule für Technik, Wirtschaft und Kultur Leipzig, Humboldt-Universität zu Berlin, Niedersächsische Staats- und Universitätsbibliothek Göttingen, Staatliche Akademie der Bildenden Künste Stuttgart und Technische Universität Wien

8 http://www.nestor.sub.uni-goettingen.de/education/nestor_e-Tutorials.pdf

9 http://www.nestor.sub.uni-goettingen.de/school_2012/index.php

10 <http://www.digcur-education.org/>

- als Innovationslabor und Think-Tank innerhalb von Bibliothek und Hochschule fungiere
- Kontakte zu nationalen und internationalen F&E-Communities herstelle
- Beratung anbiete und sich in die Lehre einbringe.

So diene sie in vielerlei Hinsicht dem Transfer von Know-how, auch und gerade in ihr bibliothekarisches und universitäres Umfeld hinein.



Abb. 4 Prof. Dr. Achim Oßwald,
FH Köln

Oßwald kam dann auf die Bibliotheken und ihr Selbstbild zu sprechen. Diese sähen sich selbst sehr positiv und betrachteten sich nicht nur als die am stärksten genutzte Kultur- und Bildungseinrichtung, sondern auch als Ort des freien Zugangs zu Wissen und Information. Die Wirklichkeit weiche davon allerdings gelegentlich ab. Zwar seien die Lesesäle voll, doch wisse man nicht genau, ob dies der Attraktivität der Bibliothek als Lernort oder nur dem Phänomen Massenuniversität geschuldet sei. Bibliotheken griffen Innovationen in Bezug auf Medienangebot, Technik und Dienstleistungen auf, seien dabei aber in den Regularien und Prozessen durchweg auf Tradition bedacht. Daher würden sie von Außenstehenden oft eher als Institutionen mit vielen

Beschränkungen wahrgenommen. Zwar hielten sie ein durchaus differenziertes Medien- und Ressourcenangebot vor, doch seien viele Nutzer von Google u.Ä. der Ansicht, alle für sie relevanten Informationen und Ressourcen ließen sich im Internet finden – und das auch noch kostenlos. Dabei machten sie sich nicht klar, dass es größtenteils Bibliotheken seien, die solche Informationen zuvor bereitgestellt hätten. Vor diesem Hintergrund würden Hochschulbibliotheken – ungeachtet ihrer in aller Regel gesetzlichen und institutionellen Absicherung – inzwischen hinsichtlich ihrer Strukturen und Servicequalität sowohl von Hochschulleitungen als auch von einzelnen Nutzergruppen kritisch hinterfragt.

Mit vielen Schwierigkeiten sähen sich Bibliotheken durch die digitalen Medien konfrontiert, deren Anteil am Medienangebot kontinuierlich zunehme. Aufgrund der hohen Diversität insbesondere der Beschaffungsverfahren und der Lizenzbedingungen lasse sich für das Ressourcenmanagement nur schwer ein einheitliches Konzept bereitstellen. Zusätzliche Probleme resultierten aus der Ausrichtung traditioneller Bibliothekssoftware an Printmaterialien und der damit einhergehenden Trennung von Geschäftsprozessen nach digitalen und gedruckten Materialien. Bei den Metadaten komme erschwerend hinzu, dass in Deutschland in diesem Bereich noch nationale Formate und Regeln vorherrschten, die die Datenübernahme behinderten. Zudem seien digitale Ressourcen im Katalog oft nur nachgewiesen, während die Nutzer auch schon die Möglichkeiten eines sofortigen Zugriffs erwarteten. Hier gebe es noch keine perfekten (Software-) Lösungen, aber immerhin zahlreiche ‚workarounds‘. Sie seien zwar nur suboptimal, aber besser als nichts! In einigen Bereichen verschaffe Outsourcing gewisse Arbeitserleichterungen, so bei der regalfertigen Lieferung von Büchern und Zeitschriften, bei ‚approval plans‘ und ‚patron driven acquisition‘, bei Portallösungen für digitale Ressourcen und Suchindizes für ‚discovery services‘.

Als aktuelle Handlungsfelder wissenschaftlicher Bibliotheken nannte Obwald:

- *Open Access*: Hier sei bereits einiges erreicht worden. So habe man institutionelle Repositorien eingerichtet, Policies und organisationsinterne Unterstützungsstrategien etabliert, Hochschulverlage gegründet und die Mitarbeit bei Open-Access-Zeitschriften und -Publikationen forciert.
- *Digitalisierung*: Entsprechende Projekte seien durchgeführt, Digitalisierungszentren gegründet und erste Virtuelle Forschungsumgebungen eingerichtet worden, die es weiter zu fördern gelte.
- *Forschungsdatenmanagement*: Die Sicherung, Archivierung und Publikation von Forschungsdaten trage wesentlich zur Nachhaltigkeit von Forschung und Entwicklung bei.
- *Technische und konzeptionelle Öffnung für mobile Endgeräte*: Bibliotheken arbeiteten daran, dafür entsprechende Dienstleistungen bereitzustellen – eine keineswegs triviale Aufgabe, da die gängige Software dafür oft nicht geeignet sei.
- *Cloud-basierte Bibliothekssysteme*: Bibliothekssystemhersteller entwickelten momentan solche Systeme, die Prozesse verschlanken und verbundübergreifende Katalogdaten in herstellerbezogenen Datensilos bereitstellen sollten.

Angesichts dieser neuen Entwicklungen müssten sich wissenschaftliche Bibliotheken auf ihre ursprüngliche Zielsetzung zurückbesinnen, Informationsressourcen proaktiv bereitzustellen. Zudem gelte es, die Nutzerinnen und Nutzer bei der Suche, Auswahl und Nutzung von Informationsressourcen zu unterstützen sowie Hilfestellung bei Publikationsprozessen zu leisten. Nur so könnten Bibliotheken beweisen, dass trotz Google und Ähnlichem nach wie vor Bedarf für spezifisch bibliothekarische Dienstleistungen bestehe. Um Kapazitäten für Beratung und Unterstützung der Nutzerinnen und Nutzer zu schaffen, sei es allerdings nötig, interne Prozesse soweit wie möglich zu verschlanken. Wissenschaftliche Bibliotheken müssten zudem ihre Nutzer- und Kundenorientierung generell weiter verbessern. So sollten Bibliothekare offen und proaktiv auf die Bedürfnisse und Fragestellungen seitens der Nutzerinnen und Nutzer eingehen und weniger prozess- als ergebnisorientiert arbeiten. Vision und Ziel seien „embedded librarians“ bzw. „liaison librarians“, d.h. Bibliothekare als Kommunikatoren mit fachlicher und organisatorischer Einbindung in die Forschung und Entwicklung an der Hochschule und in die eigene Fachcommunity.

Die Abteilung für Forschung und Entwicklung könne dazu beitragen, indem sie von Zeit zu Zeit das eigene Handeln und die organisatorischen Strukturen überdenke, serviceorientierte Zielsetzungen in den Fokus rücke, prototypische Umsetzungen realisiere und die Nutzerinnen und Nutzer als Kunden in Projektgruppen begleite, unterstütze und entlaste. Bibliothekarinnen und Bibliothekare müssten sich ihrerseits fachlich und organisatorisch fort- und weiterbilden, gegenüber technischen und organisatorischen Neuerungen offener werden und die mit der Neuausrichtung bibliothekarischen Handelns verbundenen veränderten Anforderungen gegenüber Softwareherstellern und externen Dienstleistern präzise artikulieren. Konsequenter Know-how-Transfer sei das Stichwort!

Wenn eine an diesem neuen Konzept orientierte Hochschulbibliothek künftig nicht nur Medien, sondern auch spezifische Services für Forschende und Studierende bereitstellen wolle, so setze dies voraus, dass die Bibliothekarinnen und Bibliothekare in den betreffenden Forschungsgruppen und Fakultäten auf der Basis persönlicher Bekanntschaft als deren Beraterinnen und Berater wirkten. Durch eine solche Neupositionierung werde die Bibliothek für viele Nutzergruppen an der Hochschule unverzichtbar!

IV.

In dem von Uwe Schwiegelshohn moderierten zweiten Tagungsabschnitt befasste sich zunächst Andreas Rauber (Technische Universität Wien) mit der Frage: „Welcher Änderungsbedarf ergibt sich bei den Service-Infrastruktureinrichtungen?“. Es folgten drei Impulsreferate mit anschließender Diskussion zur Leitfrage: „Wie verändert sich das wissenschaftliche Arbeiten?“. Antworten darauf gaben Arnulf Quadt (Universität Göttingen) für die Naturwissenschaften, Anke Holler (Universität Göttingen) für die Geisteswissenschaften und Tobias Blanke (King’s College London) für die Digital Humanities.



Abb. 5 Prof. Dr. Uwe Schwiegelshohn, TU Dortmund

Bei seiner Auseinandersetzung mit der Frage „Welcher Änderungsbedarf ergibt sich bei den Service-Infrastruktureinrichtungen?“ orientierte sich Andreas Rauber am Konzept des „Fourth Paradigm“ und zeigte auf, was daraus nicht nur für diese, sondern auch für Forscherinnen und Forscher sowie für Forschungseinrichtungen folgt.

Das auf einen von Jim Gray gehaltenen Vortrag (Gray 2007) zurückgehende und dann von Tony Hey, Stewart Tansley und Kristin Tolle (2009) im gleichnamigen Buch dargelegte Konzept des „Fourth Paradigm“ zielt auf

den Entwicklungsprozess der Forschung, der sich danach in vier aufeinander aufbauenden Stufen vollziehe. So sei auf die vor ca. 1000 Jahren entstandene, auf der Beschreibung beobachteter Phänomene beruhende empirische Wissenschaft vor ca. 100 Jahren als zweite Phase die theoretische Wissenschaft gefolgt, die aus Beobachtungen Modelle ableite. Vor ca. zehn Jahren sei dann die computationale Wissenschaft aufgekommen, die mit Rechenmodellen und Simulationen komplexer Phänomene arbeite. Gegenwärtig trete die Wissenschaft in ihre vierte Phase ein, die datenintensive Wissenschaft, die sich auch als ‚data driven science‘ oder E-Science beschreiben lasse. Sie verbinde Theorie mit Experimenten und Simulationen, führe dazu riesige Datenmengen aus verschiedenen Anwendungsdomänen und Quellen zusammen und analysiere diese mithilfe von Computerprogrammen. Geradezu dramatisch ändere sich damit auch die Art und Weise der Erkenntnisgewinnung, was sich wiederum auf die Forschungsabläufe und damit auch auf die Anforderungen auswirke, die an Service-Infrastruktureinrichtungen gestellt würden.

Was die Forscherinnen und Forscher angehe, so werde die Entwicklung von Hypothesen und Theorien, die sich früher im eigenen Kopf vollzogen habe, nun zunehmend von Computern übernommen. Da der Wert der Forschung wesentlich von den Daten und ihrer Qualität abhängen, steige auch deren Wert als Grundlage der Forschung. Dabei werde diese heute nicht mehr nur mit einzelnen Daten betrieben, sondern mit vernetzten Datenbeständen aus verschiedenen und oft über lange Zeiträume zusammengestellten Quellen. Dies erfordere ganz neue Prozesse des Data Sharing und der Data Integration. Da vernetzte Daten eine ungeheure Komplexität sowohl in technischer als auch in kognitiver und konzeptioneller Hinsicht aufwiesen, werde auch das Suchen, Finden, Verstehen, Transformieren und Nutzen der Daten immer komplexer. Umso mehr seien die Forschenden dabei auf Unterstützung durch die Forschungseinrichtungen angewiesen.

Auf die Forschungseinrichtungen kämen mit den wachsenden Datenbeständen und deren steigendem Wert nicht nur höhere Investitionen zu; angesichts der Größe der Datenvolumina, der Heterogenität der Daten und ihrer Verteilung auf unterschiedliche Einrichtungen bedürfe es auch einer stärkeren Kooperation. Da viele Daten erst durch Akkumulation über lange Zeiträume hinweg Aussagekraft gewännen, falle der Data Curation mittlerweile enorme Bedeutung zu. Verwaltung, Bewahrung und Verfügbarmachung von Daten werde zur zentralen Aufgabe von Forschungsinfrastruktureinrichtungen.

Für die Infrastruktureinrichtungen wirkten sich diese Entwicklungen dahingehend aus, dass sie nicht mehr nur Supportleistungen erbrächten, sondern nun selber im Zentrum der Forschung stünden. Die Bewältigung großer Volumina von Daten teilweise höchst heterogener Art stelle sie vor große Herausforderungen. Dabei sei auch die Komplexität fachspezifischer Anforderungen enorm hoch. Das entsprechende Problembewusstsein finde sich bei den Forschenden freilich bislang oft nur gering ausgeprägt. Angesichts der teilweise gänzlich neuen Aufgabenstellungen, die wiederum ganz neue Kompetenzen erforderlich machten, stelle sich auch die Frage, ob es für die Bewältigung der damit verbundenen Probleme nicht einer völlig neuen Art von Infrastruktureinrichtungen bedürfe.

Für die Services selbst bedeute dies, dass klassische Leistungen wie Speicherung und Datenmanagement durch die enormen Volumina und die hohe Heterogenität der Daten auch in technischer Hinsicht schwieriger würden. Die mit Data Sharing, Data Transformation und Data Citation verbundenen neuen Aufgabe stellten auch die Service-Leister vor große Herausforderungen. Zu den künftigen Kernaufgaben gehöre neben Data Visualisation auch die Erarbeitung von Data Management Plans. Eine wichtige Rolle spielten in diesem Zusammenhang Qualitätssicherung, Zitierung sowie rechtliche und ethische Aspekte.

Essenziell für die Organisation sei ein neues Selbstbild und Mission Statement. Gefordert seien schlüssige Antworten auf zentrale Fragen wie: Was sind die Kernaufgaben? Was lässt sich outsourcen? Können alle Wissenschaftsdisziplinen abgedeckt werden? Bei der Speicherung werde dies sicher möglich sein, im Hinblick auf fachspezifische Aufgaben unter Umständen aber nicht. Wie solle sich die domänen- und technologiespezifische Kooperation gestalten? Auch müsse man sehen, dass die Bewahrung von Daten etwas völlig anderes sei als die Bewahrung von Büchern. Hier benötige man völlig neue Prozesse, Organisationsstrukturen, Geschäftsmodelle und Kompetenzen.

Entscheidende Bedeutung komme in diesem Zusammenhang der Aus- und Weiterbildung zu. Hier seien die Herausforderungen am größten, hier stehe man noch ganz am Anfang. Zwar werde inzwischen viel darüber diskutiert, welche Kompetenzen man benötige, doch fehle es nach wie vor an einem allgemein akzeptierten Kanon dessen, was vermittelt werden müsse. Da das Feld außerordentlich dynamisch sei, sollten Curricula nicht in Stein gemeißelt, sondern hinreichend flexibel sein. Auch Forschung und Entwicklung müsse hier vorangetrieben werden, um Best Practices zu identifizieren.

In der Ausbildung solle man sich auf die Vermittlung von Grundkompetenzen konzentrieren, an die in der späteren Weiterbildung spezifische Kenntnisse angedockt werden könnten. Auch müsse hier mehr Wert auf IT gelegt werden, sonst fehle es an der Kommunikationsbasis für den Austausch mit Partnern. Wichtig sei außerdem die internationale Vernetzung, da sich die künftigen Aufgaben nur in Kooperation über Ländergrenzen hinweg bewältigen ließen. Dies müsse auch schon in der Ausbildung zur Geltung kommen.

Generell eröffneten der radikale wissenschaftliche Umbruch und der damit insbesondere in der Forschung einhergehende, unabdingbar notwendige massive Wandel aber auch Chancen zur Erschließung neuer Zielgruppen (Forschung, Industrie, Dienstleister), zur Etablierung neuer Organisationsformen (Embedding, Integration in Fachabteilungen) und zur Erarbeitung neuer Geschäftsmodelle. Bei alledem bedürfe es internationaler Kooperation, gezielter Forschung und konsequenter Entwicklung. Auf allen drei Gebieten sei die Abteilung Forschung & Entwicklung der SUB präsent und bestens aufgestellt. Zu dieser Leistung könne man ihr nur gratulieren und wünschen, dass sie diese Position auch in den nächsten zehn Jahren zu halten vermöge.

V.

„Wie verändert sich das wissenschaftliche Arbeiten?“ – in seinem Impulsreferat befasste sich Arnulf Quadt mit dieser Frage aus Sicht der Naturwissenschaften, konkret: mit Blick auf die Teilchenphysik¹¹. Hier sei Grid Computing als Werkzeug schon seit Langem unabdingbar, um Forschung zeitgemäß betreiben zu können. Teilchenbeschleuniger und ähnliche Großgeräte allein erzeugten noch keine Erkenntnisse, man müsse die produzierten Datenmengen vielmehr auch verarbeiten und auswerten können. Art und

11 Die Teilchenphysik erforsche astrophysikalische Objekte: Galaxien, Atome und noch kleinere Teilchen. Um sie zu untersuchen, seien Instrumente notwendig: Teleskope, Mikroskope u.v.m. Die Energie, die man benötige, um bestimmte Arten von Teilchen zu entdecken, korrespondiere mit ihrer Masse. Die leichtesten Teilchen, für deren Entdeckung die geringste Energie aufgewendet werden musste, seien als erstes entdeckt und untersucht worden (in den 1960er-Jahren), die schwereren nach und nach später. Dieses Muster der ansteigenden Massen spiegele sich in der benötigten Infrastruktur wider: Wer Neues entdecken wolle, brauche heute eine Infrastruktur, die signifikant mehr Energie erzeugen könne als früher. Der Large Hadron Collider sei mit 27 km Umfang der bisher größte Teilchenbeschleuniger. Bei Experimenten erzeugten seine rund 100 Millionen digitale Auslesekanäle enorme Datenmengen.

Menge der Datenerzeugung hätten sich im Laufe der Zeit deutlich verändert. In den 1970er-Jahren habe man noch Bilder mit Analogkameras gemacht, die Objekte dabei von mehreren Seiten fotografiert und im Nachhinein versucht, 3D-Rekonstruktionen zu erstellen. In den 1990er-Jahren seien die Experimente schon komplett digital ausgelesen worden, doch hätten die Zeitspannen zwischen den Kollisionen damals noch Millisekunden betragen. Seit 2010 ließen sich mit dem Large Hadron Collider pro Sekunde Millionen Kollisionen erzeugen. Da pro Kollision Tausende von Teilchen erzeugt würden, entstünden riesige Datenmengen. Am CERN produziere man gegenwärtig jährlich 15 Petabyte Daten. Würde man sie auf DVDs speichern und diese übereinander stapeln, ergäbe das einen 20–30 km hohen Turm. Selbst bei Nutzung der Grid-Technologie zu Datenverarbeitung und -speicherung könnten Datenmengen anfallen, die sich aufgrund ihrer Größe nicht bearbeiten ließen. Daher würden komplexe Filtersysteme genutzt, mit denen sich 22.000 DVDs auf 22 reduzieren ließen.¹²



Abb. 6 Prof. Dr. Arnulf Quadt,
Universität Göttingen

In der Teilchenphysik brauche man viel Geduld. Denn das Aufspüren neuer Phänomene gleiche der Suche nach der Nadel im Heuhaufen. Dabei seien die Forscherinnen und Forscher auf eine hohe Datenmenge und auf die entsprechende Infrastruktur zu deren Verarbeitung angewiesen. In den 1950/60er-Jahren und zum Teil bis in die 1970er-Jahre seien die Daten manuell ausgewertet worden. Heute würden sie dezentral über Grid gespeichert und analysiert. In Göttingen erfolge der Zugang über das Goe-Grid, in dem verschiedene Einrichtungen zusammenarbeiteten, um Synergien zu schaffen. Daher ließen die Geräte hier auch effizienter

12 Eine wesentliche Frage sei hierbei, wie sichergestellt werden könne, dass in diesem Prozess keine wichtigen Daten verloren gingen. Dieses Thema würde allerdings den Rahmen dieses Impulsreferats sprengen.

als anderswo. Das sei allerdings nicht nur eine technische Frage, sondern habe auch mit der Ausbildung der Forscherinnen und Forscher und ihrer Kooperation zu tun. Die Ausbildung angehender Physikerinnen und Physiker dürfe sich dabei nicht auf Physik beschränken, sondern müsse interdisziplinär gestaltet sein. So habe das berühmte Higgs-Teilchen nur mithilfe statistischer Methoden gefunden werden können, die man von den Sozialwissenschaften übernommen habe.

Im Mittelpunkt der anschließenden Diskussion standen die Publikationsformen und -gewohnheiten in der Physik, wo Veröffentlichungen mehr und mehr eine große, zuweilen seitenfüllende Zahl von Autoren auswies. Da stelle sich die Frage, ob es sich hier auch um eine qualitative Veränderung handle und ob die einzelnen Autoren einander überhaupt noch persönlich bekannt seien. Dies treffe – so Quadt – in der Regel zu; schließlich arbeiteten die Beteiligten oft über Jahre gemeinsam an dem betreffenden Projekt. Signifikante Veränderungen sehe er hinsichtlich der Publikationen nicht. Allerdings habe es einen bemerkenswerten sozialen Wandel gegeben: Während in den 1960er- und 1970er-Jahren die Physik fast ausschließlich von Männern betrieben worden sei, gebe es heute unter den Forschenden auch viele Frauen. Was die Publikationen betreffe, so werde ständig diskutiert, ob unter den Autoren auch jene aufgeführt werden sollten, die an der Entwicklung der Infrastruktur mitgearbeitet hätten. Die Regel sei hier, alle Personen zu nennen, die einen signifikanten Beitrag zum Forschungsergebnis geleistet hätten. Das beantworte allerdings noch nicht die Frage, wie dieser signifikante Beitrag zu definieren sei. Ingenieure würden in der Regel mit aufgeführt – sofern sie maßgeblich zum Forschungsergebnis beigetragen hätten. Derzeit überlege man auch, ob und wie man die Karrieren von Nachwuchswissenschaftlerinnen und -wissenschaftlern auch dadurch fördern könne, dass man sie aus der Masse der Autoren gezielt heraushebe.

Anke Holler zeigte dann am Beispiel der Germanistischen Linguistik die Perspektive der Geisteswissenschaften auf. Dem Wissenschaftsrat zufolge fielen darunter insgesamt 96 Fächer. Auch wenn sie durch eine große Vielfalt der Forschungsgegenstände und Methoden geprägt seien, verbinde sie alle doch die Gemeinsamkeit der Suche nach Spuren menschlichen Denkens und Handelns. Ihren gemeinsamen Forschungsgegenstand bildeten der Mensch und seine zeichenhaften Ausdrucksformen, insbesondere seine Sprache. Es gehe ihnen allen um Sammeln, Bewahren und Interpretieren. Momentan vollziehe sich in den Geisteswissenschaften eine deutliche Hinwendung zu empi-

rischen Methoden. Am Beispiel der Linguistik ließen sich derzeit sechs signifikante Veränderungen beobachten:

1. *Theoriebildung: „Denk nicht, sondern schau!“*

Der Weg führe hier weg von introspektiven Sprecherurteilen und hin zu empirisch erhobenen Sprachdaten. Dabei entstünden auch zahlreiche Spezialkorpora zu einzelnen sprachlichen Phänomenen. Neu seien multimodale Korpora wie die „Dokumentation bedrohter Sprachen“.

2. *Ein anderer Umgang mit den Daten*

Die Linguistik habe sich von Papier und Bleistift gelöst und in Richtung Sprachtechnologie bewegt. Dabei nutze sie sprachtechnologische Werkzeuge zum Aufbau von Datensammlungen und neue Verfahren zu deren Auswertung. Angesichts dieser Entwicklung stellten sich für die Geisteswissenschaften ganz neue Fragen: nach Primär- und Sekundärdaten, nach Langzeitarchivierung und Standardisierung sowie nach dem Aufbau geeigneter Infrastrukturen, wie sie von TextGrid oder CLARIN¹³ realisiert würden.

3. *Befreiung von Irrtümern*

In vielen Wissensbereichen würden Annahmen tradiert, von denen man mithilfe quantitativer Methoden zeigen könne, dass sie längst nicht erwiesen und die entsprechenden Phänomene noch längst nicht verstanden seien. Ein schlagendes Beispiel dafür liefere das Phänomen der Präposition-Substantiv-Sequenzen, die es nach der Grammatik gar nicht geben dürfte, da in ihnen der Artikel fehle. Lange Zeit habe in der Grammatikforschung als Standardannahme gegolten, dass man es hier mit einer finiten Menge idiomatischer Strukturen zu tun habe. Mithilfe quantitativer Methoden sei dies inzwischen widerlegt.

4. *Forschungsfragen neuer Art*

Was leisten einzelne empirische Methoden? Wie können verschiedene Methoden nutzbringend kombiniert werden? Wie geht man mit konfligierenden Daten um? Welche Art sprachlichen Wissens kann überhaupt aus Korpora extrahiert werden? Fragen dieser Art stellten sich heute für die Linguistik.

5. *Vom Zettelkasten zu Google Scholar*

Durch den jederzeit möglichen direkten Zugang zu Informationen und die damit einhergehenden vielfältigen neuen Recherchemöglichkeiten beschleunige und vereinfache sich die wissenschaftliche Arbeitsweise.

13 <http://www.clarin.eu/external/>

Bisher noch zu wenig diskutiert werde in den Geisteswissenschaften die Frage nach der Auswahl der Suchergebnisse, die ja keineswegs objektiv, sondern hochgradig subjektiv geprägt sei. Welche Themen in den Fokus der Wissenschaft rückten, werde heutzutage in großem Maße durch Such- und Ranking-Algorithmen mitbestimmt – mit dem Effekt, dass sich *Mainstream-Trends* verstärkten und „Orchideen“ eher unsichtbar blieben.

6. *Teamplayer gesucht*

Die Geisteswissenschaften seien traditionell durch intellektuelle Einzelleistungen geprägt. Doch nicht jede gute Linguistin und jeder gute Linguist könne zugleich auch Experte oder Expertin in der Computerlinguistik sein. Da der Einsatz quantitativer Methoden technisches und mathematisches Verständnis voraussetze, müsse sich geisteswissenschaftliche Forschung zwangsläufig zu einer kollaborativen Forschung entwickeln. Man brauche interdisziplinäre Teams aus Fachspezialisten und Infrastruktur-Experten.

Bei alledem brauchten die Geisteswissenschaften nicht um ihre Existenz zu fürchten, wenn sie sich jetzt vermehrt an Methoden der Naturwissenschaften orientierten. Denn genuine geisteswissenschaftliche Expertise sei heute nötiger als je zuvor. So brauche man sie beispielsweise beim Aufbau von Repositorien und Archiven, deren Datenmodelle nur mit fachspezifischem Wissen entwickelt werden könnten. Und schließlich gelte: Daten sind nicht gleich Wissen. Oder anders gesagt: Erkenntnisgewinn ist ohne einen klugen Kopf zur Interpretation der Daten nicht möglich.



Abb. 7 Prof. Dr. Anke Holler,
Universität Göttingen

Im dritten Impulsreferat stellte Tobias Blanke Veränderungen im wissenschaftlichen Arbeiten aus der Perspektive der Digital Humanities dar. Eingangs unterstrich er noch einmal, welche außergewöhnliche Leistung sich mit dem zehnjährigen Bestehen der Abteilung Forschung & Entwicklung

der SUB verbinde. Diese Abteilung sei schon insofern einmalig, als in ihr mit Forschung und Infrastruktur zwei ganz unterschiedliche ‚Welten‘ aufeinanderträfen. Ein solches Konzept zeuge von der Weitsicht jener, die es damals auf den Weg gebracht hätten. Denn wenn man die großen Herausforderungen der Zukunft wie Research Data Management oder Digital Preservation bewältigen wolle, müsse man diese beiden Bereiche unbedingt zusammendenken und verbinden.



Abb. 8 Dr. Tobias Blanke,
King's College London

Die Frage nach den Veränderungen wissenschaftlichen Arbeitens müsse aus der Perspektive der Digital Humanities von zwei Seiten her gestellt werden. So gehe es um Veränderungen nicht nur der Geisteswissenschaften, sondern auch der Computer- und der Naturwissenschaften. Bei den Digital Humanities richte sich der Blick demgegenüber in aller Regel nur auf die Geisteswissenschaften und deren Übernahme von Methoden aus den Naturwissenschaften. Wohin sich die Digital Humanities künftig entwickelten, lasse sich zum gegenwärtigen Zeitpunkt schwer abschätzen. Lange Zeit seien sie eine Art Rebellenbewegung einiger Weniger gewesen, die im Computer sehr früh nicht den Feind, sondern ein nützliches Werkzeug gesehen hätten, mit dem sich auch geisteswissenschaftliche Forschung betreiben ließ. Mittlerweile sei der Computer in der Mitte der Geisteswissenschaften angekommen. So gebe es eine ganze Reihe höchst produktiver Partnerschaften mit den Natur- und den Computerwissenschaften. Auch nehme man in der Wissenschaft inzwischen zur Kenntnis, dass auch die Geisteswissenschaften über eine Fülle von Daten verfügten. Sogar *Nature* habe 2010 ein Feature ver-

öffentlich, das sich mit „Big Data in den Geisteswissenschaften“¹⁴ beschäftige. Insoweit seien die Digital Humanities im Mainstream angekommen.

Wie bei anderen interdisziplinär ausgerichteten Wissenschaften sei es auch im Fall der Digital Humanities nicht einfach, zu einer allseits akzeptierten Definition und einem einheitlichen Verständnis dessen zu gelangen, was sie ausmache. Bei aller momentanen Popularität gehe ihnen noch immer der Ruf voraus, nicht mehr als eine moderne Hilfswissenschaft zu sein. Umso mehr müsse den Digital Humanities daran gelegen sein, ihr eigenes Forschungsinteresse stärker herauszuarbeiten, eigene genuine Fragestellungen zu entwickeln und dazu überzeugende Publikationen vorzulegen. Ansätze dazu gebe es bereits. Ein Beispiel dafür sei Distant Reading. Ein anderes sich entwickelndes Forschungsprogramm, das die Digital Humanities als eigenständige Disziplin auszuweisen vermöchte, beschäftige sich mit der Analyse von Big Data. Fähigkeiten und Mehrwert der Geisteswissenschaften könnten sich hier beim Identifizieren von Verbindungen und Entwickeln von Stories aus den Daten erweisen. In der historischen Forschung sei dies essenziell, um Daten zu ihrem Entstehungskontext in Beziehung zu setzen. Es gebe also viel zu tun – und wenn die Digital Humanities es nicht selber anpackten, besetzten Andere das Feld.

In der anschließenden, alle drei Impulsreferate einbeziehenden Diskussion zeigte sich Norbert Lossau geradezu begeistert von der Selbstverständlichkeit, mit der Arnulf Quadt über seine Instrumente gesprochen habe. In den Digital Humanities werde unablässig gefragt, wo denn der Mehrwert dieser Disziplin liege, und damit ständiger Rechtfertigungsdruck erzeugt. Im Grunde sei dies völlig überflüssig. Wenn viele eher traditionell orientierte Geisteswissenschaftlerinnen und Geisteswissenschaftler geradezu pikiert darüber schienen, dass es jetzt neue Fördertöpfe gebe und sich eine gewisse Annäherung dieser Fächer an andere Disziplinen vollziehe, komme in dieser Abwehrhaltung viel Selbstverletztheit zum Ausdruck. Dabei könnten ja auch diese Wissenschaftlerinnen und Wissenschaftler ganz selbstverständlich mit den neuen Methoden arbeiten.

Dem stimmte Tobias Blanke zu. Die herkömmliche Infrastruktur der Universitäten – die Bibliotheken – würden mehrheitlich von Geisteswissenschaftlerinnen und Geisteswissenschaftlern genutzt. Dass man gegenüber den neuen Methoden und Verfahren Vorbehalte habe und sie teilweise als eher

14 *Nature Reviews Genetics* 11 (September 2010), pp. 647–657, <http://www.nature.com/nrg/journal/v11/n9/full/nrg2857.html>

problematisch ansehe, habe nicht zuletzt mit der traditionellen geisteswissenschaftlichen Forschung und der dort vorherrschenden Arbeitsweise zu tun. Dabei verkenne man jedoch, dass die Arbeit mit digitaler Methodik hohen Erkenntnisgewinn erzeuge. Umso mehr komme es darauf an, dass man entsprechende Resultate der Digital Humanities publiziere.

Anke Holler sah hier vor allem die Ausbildung gefordert. Sie selbst habe nicht nur Linguistik, sondern auch Psychologie und Informatik studiert. Doch eine solche Fächerkombination sei in den Geisteswissenschaften die Ausnahme. Der Großteil der geisteswissenschaftlichen Studierenden zeige nach wie vor wenig Interesse an Mathematik und naturwissenschaftlichen Methoden. Dementsprechend könne der zur Nutzung der neuen Methoden notwendige mathematisch-technische Hintergrund nicht als selbstverständlich vorausgesetzt werden. Diese Lücke gelte es zu füllen. Für eine Infrastruktureinrichtung liege hier eine besondere Chance.

Arnulf Quadt erschien die Integration von Statistik und Informatik in den Lehrplan auch in den Naturwissenschaften noch keineswegs als Regelfall. So gehörten beispielsweise in der Teilchenphysik an der Universität Göttingen statistische Methoden lange Zeit nicht zum Curriculum. Mittlerweile sei das anders, aber es habe eben gedauert. Gleiches gelte für die Instrumentierung, das Elektronikpraktikum und die Computerausbildung. Hier arbeite man mit der SUB und der Universitätsmedizin zusammen. Für die Studierenden und ihre Ausbildung sei das sehr gut, doch werde es keineswegs allgemein akzeptiert und auch von der Fakultät nicht auf breiter Basis mitgetragen.

VI.

Den Abschluss des Symposiums bildete die von Axel Horstmann moderierte Podiumsdiskussion, die sich mit der Frage befasste: „Kooperationen zwischen Infrastruktureinrichtungen und Fachwissenschaft – quo vadis?“. Teilnehmer waren Ulrich Meyer-Doeringhaus (Perspektive Hochschulrektorenkonferenz), Norbert Lossau (Perspektive Infrastruktur Bibliothek), Werner Wegstein von der Universität Würzburg (Perspektive Geisteswissenschaften) und Bernadette Fritsch vom Alfred-Wegener-Institut für Polar- und Meeresforschung (Perspektive Naturwissenschaften).

Als gemeinsamer Tenor der Impulsreferate zeichne sich – so Axel Horstmann einleitend – offenkundig die Anerkennung sowohl der enormen Bedeutung der Forschungsinfrastrukturen als auch der Notwendigkeit einer engen

Kooperation zwischen ihnen und der Fachwissenschaft ab. Vor diesem Hintergrund sei es umso wichtiger zu erfahren, wo, wie und von wem denn auf diesem Feld konkret kooperiert werde.



Abb. 9 Abschlussdiskussion – Prof. Dr. Werner Wegstein, Dr. Bernadette Fritzsch, Prof. Dr. Norbert Lossau, Dr. Ulrich Meyer-Doerpinghaus, Prof. Dr. Axel Horstmann

Norbert Lossau zufolge ist beispielsweise die SUB in eine Vielzahl solcher Kooperationsaktivitäten eingebunden. So gebe es eine enge Partnerschaft zwischen Bibliothek und Fachwissenschaft etwa bei den Spezialsammlungen, wo die dort arbeitenden Wissenschaftlerinnen und Wissenschaftler auf die „key enabler“ angewiesen seien, die ihnen die Daten zur Verfügung stellten und beratend zur Seite stünden. Daraus folge, dass auch die betreffende Bibliothek bzw. deren Mitarbeiterinnen und Mitarbeiter in den späteren Publikationen genannt werden müssten, denn diese hätten wesentlichen Anteil am Zustandekommen der Forschungsergebnisse. Weitere enge Partnerschaften bestünden schon seit Längerem bei den Digitalisierungsprojekten. In der Regel träten hier die Geisteswissenschaftlerinnen und Geisteswissenschaftler mit einer interessanten Sammlung und der Bitte an die Bibliothek heran, diese digitalisieren zu lassen. Auf diesem Gebiet habe die SUB in den letzten Jahren viel geleistet. Open-Access-Publikationen seien als ein weite-

res Kooperationsfeld zu nennen. Traditionellerweise liege der Schwerpunkt auf Kooperationen mit den Geisteswissenschaften, doch arbeite die SUB zunehmend auch mit Partnern aus den Natur- und Lebenswissenschaften zusammen.

Werner Wegstein vertrat in diesem Zusammenhang die Auffassung, dass sich die Geisteswissenschaften früher oder später in ihrer Gesamtheit zu digitalen Geisteswissenschaften entwickeln und sich die Digital Humanities damit über kurz oder lang als eigene Disziplin überflüssig machen würden. Dabei dürfe man nicht vergessen, dass sich die Daten der Natur- und Lebenswissenschaften von denen der Geisteswissenschaften qualitativ unterscheiden. Während die Bibliotheken bei den natur- und lebenswissenschaftlichen Daten lediglich eine vermittelnde Rolle spielten, beherbergten sie im Falle der Geisteswissenschaften die Quellen der Forschung selbst. Was die Linguistik betreffe, seien Veränderungen im Umgang mit ihren Quellen nicht zu übersehen. Habe man früher schwer oder gar nicht auffindbare Quellen – darunter vielleicht auch solche, an denen man möglicherweise nicht sonderlich interessiert war – gar nicht in den Forschungsfokus einbezogen, gebe es heute in der Korpuslinguistik ‚kein Entrinnen‘ mehr. Und je mehr man mit Korpuslinguistik arbeite, desto größere Überraschungen seien für die Theoriebildung zu erwarten.

Bernadette Fritsch verwies für die Klimawissenschaften in diesem Zusammenhang auf das seit einigen Jahren bestehende System der Weltzentren. Hier finde sich möglicherweise schon etwas von dem aufgegriffen, was Andreas Rauber in seinem Vortrag gefordert habe; dass es nämlich neben den fraglos wichtigen Bibliotheken weitere Infrastruktur-Serviceeinrichtungen mit spezieller Ausrichtung auf Forschungsdatenmanagement geben sollte. Solche Datenzentren seien in den Klimawissenschaften anerkannte Einrichtungen. Und die Kooperation zwischen ihnen und der Wissenschaft sei in der Community tief verankert.

Die Frage, was diese Kooperationen vorantreibe und woher die entscheidenden Impulse dafür kämen, lasse sich – so Ulrich Meyer-Doeringhaus – nicht leicht beantworten. Es bleibe zu hoffen, dass die Anstöße aus der Forschung kämen. Würden die Kooperationen einseitig von der IT-Seite vorangetrieben, könnte dies am Ende womöglich nur zu einem großen, nicht mehr zu bewältigenden und insofern sinnlosen Datenmeer führen. Noch bedenklicher wäre es allerdings, wenn die Impulse hauptsächlich von den Forschungsförderorganisationen kämen. Demgegenüber unterstrich Norbert Lossau, dass erfolgreiche Kooperationen immer getragen würden vom gleich-

mäßigen Zusammenspiel unterschiedlicher Akteure, im Idealfall von Teams aus Vertretern der Fachwissenschaft und des Infrastruktur-Bereichs. Was die einschlägigen Förderprogramme betreffe, so lasse sich hier seitens der Wissenschaft durchaus steuernd eingreifen.

Blockaden resultieren nach Einschätzung von Werner Wegstein insbesondere aus den urheberrechtlichen Regelungen. So ließen sich Wissenschaftlerinnen und Wissenschaftler oft nur ungern oder gar nicht in die eigenen Zettelkästen schauen. Hinzu komme, dass die Normierung der Daten noch nicht weit genug vorangetrieben worden sei. Ohne Vergleichbarkeit der Datengrundlagen aber lasse sich nicht kollaborativ arbeiten. Er sei daher ein glühender Anhänger der Text Encoding Initiatives (TEI)¹⁵, denn sie schafften die notwendigen Voraussetzungen für Vergleiche. In dieser Richtung müsse und werde man weitergehen. Studierende seien durchaus interessiert, mit solchen neuen Methoden zu arbeiten. Denn im Gegensatz zu interpretativen Arbeiten bestehe hier die Chance, am Ende klare Aussagen im Sinne von „richtig“ oder „falsch“ treffen zu können.

Mit Blick auf den internationalen Wettbewerbsdruck unterstrich Bernadette Fritsch für die Klimawissenschaften, dass enge Verbindungen zu internationalen Projekten für diese Fächer angesichts ihrer Einbettung in die internationale ‚scientific community‘ ebenso selbstverständlich wie unverzichtbar seien.

Vergleichsweise zurückhaltend äußerten sich die Podiumsteilnehmer zu der Frage, ob von den in Rede stehenden Kooperationen zwischen Fachwissenschaft und Infrastruktureinrichtungen eine substanzielle Leistungs- und Qualitätssteigerung auf Seiten der Wissenschaft zu erwarten sei. Ulrich Meyer-Doeringhaus diagnostizierte in einschlägigen Statements einen gewissen ‚Datenoptimismus‘. Dabei werde seinem Eindruck nach nicht scharf genug getrennt zwischen den Daten selbst, die an sich eigentlich ‚sinnlos‘ seien, und den Menschen, die diese Daten erforschten, sie zu interpretieren und damit überhaupt erst mit Sinn zu erfüllen vermöchten. Werner Wegstein verwies in diesem Zusammenhang auf neue Möglichkeiten der Linguistik, in ihren Korpora vorhandene, bis dato jedoch unentdeckt gebliebene Fehler durch „crowd correcting“ aufzuspüren und zu beseitigen.

Deutlich positiver schätzten die Podiumsteilnehmer den wissenschaftlichen Ertrag der mit den neuen Verfahren einhergehenden Empirisierung und

15 <http://www.tei-c.org/index.xml>

der Erschließung großer Datenmengen ein. So vertrat Tobias Blanke die Auffassung, dass sich damit neue Fragestellungen entwickeln ließen. Ein Beispiel liefere das von ihm mitbearbeitete Projekt „European Holocaust Research Infrastructure (EHRI)“¹⁶, welches Daten aus vielen verschiedenen Quellen zusammenbringe. Jetzt gebe es hier zum ersten Mal eine Datenbank mit über einer Million Einträgen, die sich auf die Opfer des Holocaust bezögen. Das ermögliche eine Perspektive ‚von unten‘, d.h. aus der Sicht einzelner ‚einfacher‘ Individuen (im Gegensatz zur Ebene der Herrschenden oder anderer herausgehobener Persönlichkeiten), die man früher nicht habe darstellen können. Ein weiteres Beispiel seien die „daily records“ des Londoner Strafgerichtshofes aus dem 18. Jahrhundert, die wertvolle Kontextinformationen zu den einzelnen Straffällen lieferten und damit das alltägliche Leben im London der Zeit ins Licht rückten.

Andreas Rauber sah dies ähnlich. Die neue Art, Wissenschaft zu betreiben, müsse und werde kommen. Ein Blick auf die Naturwissenschaften zeige, dass die Produktion von Daten und Publikationen derart ausufere, dass man sie als einzelner Forscher gar nicht mehr zu überblicken vermöge. Daher brauche man, um überhaupt noch sinnvoll weiterforschen zu können, geeignete Hilfsmittel, um die Fülle an Informationen zu analysieren und Muster zu erkennen. Die neuen Methoden zur Datenanalyse und Hypothesengenerierung würden sich mit Sicherheit durchsetzen.

Auch Heike Neuroth zeigte sich überzeugt, dass die Erweiterung des Datenvolumens neue Fragestellungen generieren werde. Es komme allerdings wesentlich darauf an, dass man bei der Arbeit an und mit den Daten verstehe, was da vor sich gehe, was auf welche Weise untersucht werde und welcher Algorithmus sich dahinter verberge. Nur dann könne man gute Forschung machen. Dieses Bewusstsein müsse noch stärker vermittelt werden.

Unterschiedliche Antworten gab es auf die Frage, was jetzt eigentlich von wem zu tun sei, wer oder was sich verändern müsse und wer jetzt vorrangig gefordert sei – die Service-Einrichtungen, die Wissenschaftlerinnen und Wissenschaftler selbst oder wer sonst noch? Sigrun Eckelmann von der Deutschen Forschungsgemeinschaft (DFG) verwies in diesem Zusammenhang auf mehrere einschlägige, von ihr betreute Projekte. Wenn solche Vorhaben nur von einer Seite vorangetrieben würden, führe das meist zu Schwierigkeiten. Essenziell sei hier partnerschaftliche Kooperation auf Augenhöhe. Der Wille

16 <http://www.ehri-project.eu/>

dazu müsse auch und vor allem von der wissenschaftlichen Basis kommen. In dieser Hinsicht sei noch einiges zu tun.



Abb. 10 Blick ins Plenum

Uwe Schwiegelshohn richtete seine Forderung vor allem an die Politik, die richtigen Rahmenbedingungen zu schaffen. Daran aber fehle es bislang. Wenn die Wissenschaftlerinnen und Wissenschaftler neben Blockaden in den eigenen Reihen und technologischen Problemen auch noch politische Widerstände überwinden müssten, seien sie damit schlicht überfordert.

Ulrich Meyer-Doeringhaus sah vor allem die Hochschulleitungen in der Pflicht. Zwar wünschten sie sich alle möglichst viele Nobelpreise für ihre eigene Institution; doch dass solche Erfolge auch etwas mit der Ausstattung der eigenen Bibliothek zu tun haben könnten, werde nur von den Wenigsten realisiert. So müsse in den Hochschulen ganz generell das Bewusstsein dafür gestärkt werden, dass exzellente Forschung ohne eine leistungsfähige Infrastruktur nicht zu haben sei. Als größter Hemmschuh erweise sich auf der politischen Ebene freilich schon seit Langem der Föderalismus. Seine restriktiven Regelungen und Vorschriften machten wissenschaftliche Kooperationen zwischen Institutionen aus unterschiedlichen Bundesländern extrem schwierig. Die HRK kämpfe daher für eine grundlegende Reform der föderalen Rahmenbedingungen.

Für Norbert Lossau standen die Zeichen auch in dieser Hinsicht gar nicht so schlecht. In den letzten Jahren sei das Thema Informationsinfrastruktur Gegenstand einer ganzen Reihe von Stellungnahmen und Empfehlungen gewesen, die erkennen ließen, dass die Sensibilität hierfür wachse. Dass es sich gleichwohl um einen beschwerlichen Weg handle, spüre man auch in der eigenen Hochschule – die Richtung aber sei klar. Mit dem geplanten Rat für Informationsinfrastrukturen eröffne sich eine Erfolg versprechende Möglichkeit, das Thema auf der politischen Ebene zu verankern und den Entscheidungsträgern deutlich zu machen: Hier gibt es Bedarf!

Was die Kooperation zwischen Fachwissenschaft und Informationsinfrastruktur im internationalen Vergleich betrifft, erschien Bernadette Fritzsich die deutsche Klimawissenschaft und ihre Community als gut aufgestellt. Momentan befinde sich gerade der nächste Sachstandsbericht für den IPCC¹⁷ in Vorbereitung. Deutschland sei dabei ein gewichtiger Partner, zumal es über eines der Kernzentren in diesem Datenverbund verfüge.

Mit Blick auf die Sprachwissenschaften und die Geisteswissenschaften insgesamt schätzte Werner Wegstein die Lage hier etwas weniger positiv ein. Generell müsse man wohl zugeben, dass jeder letztlich die Infrastruktur erhalte, die er verdiene. So sei es weniger eine Sache objektiver Zwänge, wenn die strukturellen Rahmenbedingungen unzureichend blieben; zumeist fehle es dann auch am nötigen eigenen Wollen.

Demgegenüber unterstrich Elmar Mittler, dass sich Deutschland im internationalen Vergleich auf diesem Gebiet nicht zu verstecken brauche. All diejenigen, die beispielsweise in Göttingen gearbeitet hätten und inzwischen im Ausland bzw. an anderen Universitäten seien, würden das sicher bestätigen. Für diese Leistung sei er persönlich sehr dankbar. Früher habe er sich gewünscht, dass eines Tages hierzulande entwickelte Software ins Ausland exportiert werde. Das sei mittlerweile mit TextGrid geschehen. Man stehe also „verdammst noch mal nicht schlecht da“!

Auf die abschließende Frage des Moderators, was sich jeder der Podiumsteilnehmer wünschen würde, wenn er einen Wunsch frei hätte, antworteten

- Bernadette Fritzsich: Impulse in Richtung einer mehrgleisigen Ausbildung und von den Wissenschaftlerinnen und Wissenschaftlern bisweilen etwas mehr Mut, auch einmal neue Wege zu gehen;

17 <http://www.ipcc.ch/>

- Werner Wegstein: noch mehr Engagement der Bibliotheken in Bezug auf Normdaten und am besten eine flächendeckende Kodierung in TEI;
- Ulrich Meyer-Doeringhaus: Göttingen sollte überall sein!
- Norbert Lossau: fünf zusätzliche entfristete Stellen in der Abteilung Forschung & Entwicklung der Niedersächsischen Staats- und Universitätsbibliothek.

Mit dem Dank an die Podiumsteilnehmer für die engagierte Diskussion und an das Publikum für anregende Beiträge und geduldige Aufmerksamkeit schloss Axel Horstmann die Veranstaltung, gratulierte der Abteilung Forschung & Entwicklung für ihre großartige Arbeit und wünschte ihr auch weiterhin nachhaltigen Erfolg.

VII.

Wie eingangs schon festgehalten: Es war ein konzeptionell wie organisatorisch rundum gelungenes Symposium, das über das Wirken der Abteilung Forschung & Entwicklung hinaus einen breiten Überblick bot über die Evolution der Informationsinfrastruktur, über die mit den neuen technischen Möglichkeiten verbundenen Herausforderungen und Chancen, über die Veränderungen, die sich damit für die wissenschaftliche Arbeit, für Forscherinnen und Forscher, Infrastruktureinrichtungen und Forschungsumgebungen, Bibliotheken und virtuelle Datenverbünde, Hochschulen und Hochschulverwaltungen, Wissenschaftsorganisationen und Wissenschaftspolitik ergeben.

Diese vielfältigen Veränderungen vollziehen sich offenkundig mit wachsender Geschwindigkeit und erfassen zunehmend auch die Geisteswissenschaften. Digital Humanities sind – das hat sich in Göttingen deutlich gezeigt – kein Randphänomen mehr, sondern auf dem Weg ins Zentrum dieser Fächer. Ob und inwieweit sie sich gegenüber nach wie vor vorhandenen Vorbehalten und Widerständen eher traditionell orientierter Geisteswissenschaftlerinnen und Geisteswissenschaftler durchzusetzen und zu behaupten vermögen, wird wesentlich davon abhängen, ob und inwieweit die ‚scientific community‘ davon überzeugt werden kann, dass sich mit den neuen technischen Möglichkeiten nicht nur alte Fragen leichter und vielleicht auch umfassender beantworten lassen, sondern sich auch ganz neue Themenfelder, Fragestellungen und Erkenntnischancen eröffnen. Die Zeichen – auch das hat das Göttinger Symposium deutlich gemacht – stehen dafür nicht schlecht.

„Empirisierung“ könnte sich hier zu einem der Schlüsselbegriffe im Blick auf künftige geisteswissenschaftliche Standards entwickeln.

Informationsinfrastruktureinrichtungen und ihre Leistungen werden im und für den Forschungsprozess künftig weiter an Bedeutung gewinnen und – so steht zu hoffen – dementsprechend auch an Sichtbarkeit und Wertschätzung innerhalb von Wissenschaft und Wissenschaftsadministration. In Göttingen war man sich darin jedenfalls weitgehend einig. Das galt auch für die Forderung, das Zusammenspiel von Forschung und Forschungsinfrastrukturen – seien dies nun Bibliotheken, Rechenzentren oder Virtuelle Forschungs-umgebungen – weiter zu verbessern, die Kooperationen zu verbreitern und zugleich enger zu gestalten. Partnerschaft auf Augenhöhe war hier das Stichwort. Für die Bibliotheken liegen darin Herausforderung und Chance zugleich.

Kooperationen dieser Art sind nicht voraussetzungslos. Auch darin bestand in Göttingen breiter Konsens. Sie können nur gelingen, wenn auch die Forscherinnen und Forscher selbst über die nötige „Informationskompetenz“ verfügen. Das betrifft insbesondere die Geisteswissenschaften. Hier – aber nicht nur hier – muss man nach allgemeiner Überzeugung bereits in der universitären Lehre und Ausbildung ansetzen und entsprechende Module in die Fachcurricula integrieren.

Damit kommen die Hochschulen und ihre Leitungsebenen ins Spiel. Sie sind an dieser Stelle besonders gefordert. Ähnliches gilt mit Blick auf die akademische Anerkennung und angemessene Honorierung von Leistungen, die im Infrastrukturbereich erbracht werden. Dabei müssen auch die Forschungsförderorganisationen wirksame Impulse in dieser Richtung setzen und ihre Fördermöglichkeiten gegebenenfalls entsprechend erweitern.

Außer Frage stand bei den Teilnehmern des Symposiums schließlich auch, dass es nachhaltige Fortschritte auf diesem Gebiet nur geben kann und wird, wenn Wissenschaftspolitik und -verwaltung die notwendigen Rahmenbedingungen schaffen – sowohl in finanzieller Hinsicht als auch mit Blick auf administrative Regelungen, die die Zusammenarbeit zwischen den Beteiligten erleichtern. Dass der Föderalismus mit seinen oft nur schwer nachvollziehbaren Restriktionen und Barrieren dabei als eines der Haupthindernisse ausgemacht wurde, konnte auch in Göttingen nicht überraschen.

Und – last but not least – der Anlass selbst: zehn Jahre Abteilung Forschung & Entwicklung der SUB Göttingen! Das beispielhafte Wirken dieser Abteilung bildete in vielen Tagungsbeiträgen den durchgängig positiven Referenzpunkt. So wurden ihre Verdienste – und das ganz zu Recht – immer

wieder mit Nachdruck hervorgehoben. Denn dass Deutschland auf diesem Gebiet eine führende Stellung einnimmt, ist auch und gerade dieser Abteilung zu verdanken. Und wenn es noch eines Beweises für das Gelingen einer Kooperation zwischen Bibliothek und Fachwissenschaft bedurfte, so konnten ihn die Symposiumsteilnehmer vor Ort direkt in Augenschein nehmen.

Auch die Autoren dieses Berichts möchten daher nicht schließen ohne ihre ganz persönliche Gratulation an diese Abteilung und die allerbesten Wünsche für viele weitere Jahre wegweisender erfolgreicher Arbeit.

Literaturverzeichnis

- Gray, Jim (2007): eScience – A Transformed Scientific Method, http://research.microsoft.com/en-us/um/people/gray/talks/NRC-CSTB_eScience.ppt; Transkript: „Jim Gray on eScience: A Transformed Scientific Method“, http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_jim_gray_transcript.pdf.
- Hey, Tony; Tansley, Stewart; Tolle, Kristin (Hrsg.) (2009): *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond: Microsoft Research.
- HRK (2012): Hochschule im digitalen Zeitalter: Informationskompetenz neu begreifen – Prozesse anders steuern, Entschließung der 13. Mitgliederversammlung vom 20.11.2012. <http://goo.gl/tCRoK>.
- Neuroth, Heike; Oßwald, Achim; Scheffel, Regine; Strathmann, Stefan; Jehn, Matthias (Hrsg.) (2009): *nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung: Version 2.0*. Boizenburg: Hülsbusch.
- Neuroth, Heike; Oßwald, Achim; Scheffel, Regine; Strathmann, Stefan; Huth, Karsten (Hrsg.) (2010): *nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung: Version 2.3*. http://nestor.sub.uni-goettingen.de/handbuch/nestor-handbuch_23.pdf.
- Neuroth, Heike; Strathmann, Stefan; Oßwald, Achim; Scheffel, Regine; Klump, Jens; Ludwig, Jens (2012): *Langzeitarchivierung von Forschungsdaten – eine Bestandsaufnahme*. Boizenburg: Hülsbusch. Online: <http://nestor.sub.uni-goettingen.de/bestandsaufnahme>.

Forschungsprojekte

Programming for the Future?

The Specification of User Requirements in the Design of Virtual Research Environments for the Arts and Humanities

Kathleen M. Smith & Fotis Jannidis

Introduction

Many people involved in the Digital Humanities believe that Virtual Research Environments (VREs) offering integrated access to research data and the tools to analyze and work with them will be an important part of how scholars in the humanities will work in the future. Based on this assumption, there has been a proliferation of VREs within the last decade; in the meantime, it has become clear that VREs as technical and social constructs are themselves research objects. There are expanding numbers of digital projects and initiatives dealing with humanities research and data.¹ The following article concentrates on one aspect of VREs in the humanities: the question of how projects determine their user requirements. The increasing involvement of additional partners and stakeholders, such as information specialists, technical specialists, librarians, archivists, and software developers, is seen as essential for the success of these initiatives. However, the primary source of input and information about the requirements for new tools and infrastructures must remain the researchers themselves, for whom these VREs are intended.

Developing complex systems is a challenging task;² developing digital tools and services suited to humanities researchers on a broader scale is especially difficult, however, for two main reasons that could be called the translation problem and the time lag problem. The problem of translation arises from the heterogeneity of the disciplines involved. Developers and program-

1 The JISC VRE Knowledge Base provides an overview of projects and literature; see <http://misc.jisc.ac.uk/vre/>.

2 User requirements are often a central problem; see the impressive list in Sutcliffe (2002: 2 f.).

mers, as well as information specialists such as librarians, archivists, and information infrastructure specialists, come from very different contexts and use field-specific technical languages that differ from those of humanities researchers. Creating a list of what users want and need requires in-depth knowledge of the user community and the research context, since much information will be considered self-evident by those within that community when communicating with external technical experts. Translating between these different worlds is a time-intensive activity, and the problem with many attempts to collect an accurate understanding of user needs is this language barrier and the corresponding potential for misunderstanding.

The problem of time lag can be found in all disciplines adapting new technologies, but it may be more extreme in the humanities. The arts and humanities have been slower to integrate computer-supported methods of analysis into traditional research processes for a variety of reasons, including the heterogeneity of humanities data and analytical approaches, the lack of existing resources catering to humanities researchers, and the lack of financial and technical support for developing new resources on an individual level. The situation in the humanities is often depicted as a type of evolutionary process: the innovators and early adopters in many disciplines are actively exploring the new possibilities of digital data and computer-based research methods, while the majority of researchers are slowly adopting a few new ways of using and analyzing data, and a small group of researchers remains uninterested in these new possibilities. This view of change simplifies the situation and is dominated by one traditional metaphor of progress: namely, an army with a vanguard, a middle section consisting of the majority of the army, and a rearguard. This metaphor is misleading because it implies that the main group will automatically follow the lead of the vanguard, but as far as we can tell from the developments in the humanities and digital humanities, the situation is much more complex. The burden of developing new tools or using new methods that require new skills can be so cumbersome that this research field is only viable for a small group that is prepared to invest the necessary time and effort. Occasionally the new research methods and approaches developed by innovators and adapted by early adopters to their needs prove unproductive and therefore end up with no users at all.³

³ Potentially, researchers who are not interested in the technological possibilities could also develop other new methods or more general practices, but this has not been the case thus far.

On the basis of the complex nature of technical innovation in the humanities, one fundamental problem of VREs becomes clear if we assume three premises that describe the usual state of affairs: a stable, production-ready VRE takes 6–10 years to develop from scratch (working within the limits of project funding), a software project needs a clear set of user requirements to determine the goal of the project (and the means of measuring its success), and the end users involved in planning and developing the VRE are mostly recruited from the group of early adopters. How can those creating a VRE make sure that their plans will meet the requirements of the larger community of researchers in 6–10 years? How can they anticipate future requirements and possible new forms of use? User requirements and needs will change, and unanticipated developments (technical as well as cultural) will drive those requirements; these remain factors that are impossible to plan and anticipate in advance. Yet all projects will need to accommodate unexpected use scenarios and new forms of analysis in order to stay relevant and useful, and the certainty that there will be unanticipated needs can be factored into project development. Therefore they will adapt specific strategies out of the vast array of possible ways to collect user requirements and keep track of them during project development (Courage & Baxter 2005; Grady 2006).

Project Approach to User Requirements

In this article, we present three case studies of VRE research projects in the humanities in order to investigate how each addresses this issue. These projects have been selected for the following reasons. They are all projects to develop VREs for researchers in the humanities, specifically for text-based disciplines within the humanities. They are also research projects funded by national or European funding agencies that consist of partners from multiple institutions, rather than a project based at one institution or university. Since this geographic dispersion requires close teamwork and coordination, these projects are required to coordinate and document their approach. They are all projects that have as explicit goals the importance of supporting communication during all stages of the project among the different communities involved in the project: the developers and programmers, the information specialists, librarians and archivists, and humanities researchers.

The term “user requirements” is used here in a very broad sense: it refers to the description of essential or ideal features and functions in relation to the

architecture in general (such as the desire to use tools inside a browser), in relation to research data (like specific metadata), independent tools (such as named entity recognition or text-image linking), the addition of features to a tool (like adding support for xml files to a collation tool), and to improvements in the user interface to make the learning curve less steep and generally improve accessibility and productivity of a VRE (like adding tool tips to all elements of the GUI). In addition to the general humanities scholar using a VRE, it is also necessary to consider the small but very important group of developers who are either adapting tools (including VREs) to a project's needs or developing new tools in the context of a VRE – independently of the group maintaining the VRE. These developers, programmers, and computer scientists are also an important source for requirements concerning their specific needs on all the levels mentioned earlier (for example, the desire to have a REST interface for the repository of the VRE).

Because VREs bring together institutions and people with different backgrounds, requirements are not limited to the technical aspects alone, but also include policies, licenses, and best practices (such as predefined open access licenses).

Obviously this wide spectrum of user requirements cannot be covered by one method alone. An advisory board, for example, is a valuable institution for making sure that the general architecture and policies of a VRE meet with broader consent, but only seldom will it result in useful feedback about what kind of tools or features are lacking. Presentations of a VRE to a general (scholarly) public will produce many suggestions about tools and features, but it is difficult to assess how representative these demands truly are. Screen recording the work sessions of users in the VRE and trying to analyze their problems is a reliable method to improve usability of a tool but may fail to deal with more general questions such as whether the tool addresses a real need. Therefore, most VREs use very different methods to research user requirements on these different levels. The TextGrid project demonstrates a broad-based approach by encouraging collaborative research and developing tools in a research environment for a wide variety of disciplines in the humanities. One of the motivations behind TextGrid was the intention of supporting interdisciplinary, collaborative research methods among researchers from various disciplines. In contrast, CENDARI and EHRI are both projects that have a narrower focus on a research topic and community within the larger disciplinary context of history, and both share the specific goal of supporting archival research processes and methods. All of these projects have

activities and work packages that are specifically oriented towards the collection and analysis of user requirements, and all have activities intended to accommodate both current needs and future requirements on the part of their primary user communities.

TextGrid⁴

TextGrid is an infrastructure consisting of a repository and an interface with tools and services for working with humanities research data. It was one of the first humanities VRE projects in Germany and began as part of the German Grid Initiative (D-Grid) funded by the German Federal Ministry of Education and Research (BMBF) (cf. Schwiegelshohn 2008: 14). The TextGrid research group consists of sixteen partners in Germany over the three phases of the project.⁵ The target users of TextGrid are arts and humanities researchers working with digital texts and image files. The first project phase (2006–2009) and the second project phase (2009–2012) ended with the release of a repository and software interface with tools for working with text-based research methods. The third project phase (2012–2015) is focused on developing a long-term operational organization to ensure continued operation and development beyond the end of the project duration.

TextGrid's initial approach to understanding user requirements relied primarily upon the development of tools in combination with researchers. The researchers involved in TextGrid collected their understanding of the user requirements on the level of tools and features in specifications for tools and use cases which were circulated to a wider audience of interested colleagues. At a very early stage, there was also a questionnaire distributed to a general audience of editors of scholarly texts at their bi-annual conference but the results proved to be rather unsatisfactory, perhaps because most of those who participated in the survey had only very limited experience with digital tools at that moment. Communication with end users was a high priority from the beginning, however, and involvement of users in the development stages was seen as the way to accomplish the project goals. Public relations and publicity were viewed as methods to involve researchers in

4 See also the contributions to this volume by Radecke/Göbel/Söring (pp. 85 ff.), and Funk/Veentjer/Vitt (pp. 277 ff.).

5 For the current partners, see the TextGrid research group at <http://www.textgrid.de/en/ueber-textgrid/forschungsverbund/>.

the project on an ongoing basis.⁶ The proposal for the first project phase (2006–2009) emphasized supporting collaborative research tasks through tool development, which took place in combination with feedback from the Fachbeirat, or the consulting committee of expert users.⁷ Technical development was planned with the understanding that the development and emphasis on collaborative tools would attract new users and user communities from the humanities disciplines, who would adapt their research methods and research analyses accordingly (cf., for example, Küster/Ludwig/Aschenbrenner 2007). The requirements concerning the data were determined by a working group consisting of computer scientists, digital humanists and librarians who defined a metadata schema that can be extended by projects but provides a basic set of information. Because researchers also saw the need to go beyond a mere full text search, a system of baseline encodings for different text types was defined, but proved to be too complex to be implemented in a reliable and scalable way.

Community-building and coaching were addressed by a specific work package in the second project phase (2009–2012).⁸ During this phase, early experiments with screen recordings of users working with tools were applied more systematically as part of a work package and a PhD student wrote a doctoral thesis on the topic of “usability of VREs.” Specific workshops were held to make developers familiar with the architecture of TextGrid and the programming paradigm, but it was soon understood that the information gained about their specific user requirements was equally important even though it was difficult to fulfill all their wishes.

The third phase, which is concerned primarily with organizational stability rather than technical development, emphasizes the importance of understanding user requirements. The proposal for the third project phase (2012–2015) stated that one of the goals for this phase is to refine the project’s concept of the requirements from the scholarly communities („[...] den erweiterten und steigenden fachwissenschaftlichen Anforderungen gemäß

6 See, for example, the report “R 6.1 Öffentlichkeitsarbeit” (Liess/Aschenbrenner/Kerzel 2006).

7 See the first phase project proposal at <http://www.textgrid.de/fileadmin/antraege/antrag-1.pdf>.

8 See the second phase project proposal at <http://www.textgrid.de/fileadmin/antraege/antrag-2.pdf>.

konzeptionell zu optimieren“).⁹ One of TextGrid’s main goals evolved over time to support the creation of a community of users who would then work together to decide what TextGrid should be based on their research processes and workflows, rather than adapting their workflows to the tools. In recognition of the importance of non-technical user requirements, TextGrid increased support for training, instruction, outreach and advising functions for researchers preparing their own proposals for research funding.

The TextGrid project deals with the issue of anticipating its users’ future needs in several ways: by supporting the establishment and use of standards in metadata and data formats, which will support future reusability; by working to ensure long-term sustainability of its tools and resources; and by encouraging researchers to develop their own tools in combination with advice and assistance from TextGrid. Since future requirements are uncertain by nature, the best way to prepare for them is to keep existing data accessible and identifiable because these data sets will be the foundation for future work. Part of keeping the data accessible involves maintaining a suitable interface for accessing the data, which is why establishing a sustainable organizational form for TextGrid became a priority. Additionally, the TextGrid interface is designed to be expandable and modular so that external developers can create new tools and services in response to user requests: it is open source software¹⁰ and the source code is freely available under the LGPL 3 (GNU Lesser General Public License).¹¹ One of the most commonly repeated requests from external users and developers was to substitute the Eclipse-based user interface with a browser interface, and this need was therefore recognized as an important factor for the future acceptance of this VRE and became part of the workload in the third phase as well. Another factor in attempting to encourage adaptation is addressed through providing support and assistance in the preparation of funding proposals for external projects wishing to use TextGrid. This option builds upon the available resources in the TextGrid VRE, such as the interface and repository, while development of future tools and services becomes solely researcher-driven.

9 See the third phase project proposal at <http://www.textgrid.de/fileadmin/antraege/antrag-3.pdf> (Section1 “Ziele,” p. 6).

10 See <http://www.textgrid.de/registrierungdownload/download-und-installation/>.

11 See <https://projects.gwdg.de/projects/tg-search/repository/revisions/master/entry/LICENSE.txt>.

The European Holocaust Research Infrastructure (EHRI)¹²

EHRI is developing a VRE to support collaborative research in archival and institutional holdings related to the Holocaust, and to reunite geographically-dispersed collections.¹³ Its project duration is four years (2010–2014) and it is funded by the European Commission as part of the Seventh Framework Programme.¹⁴ Sister projects in the Seventh Framework Programme include CENDARI and ARIADNE.¹⁵ EHRI partners include nineteen institutions, from libraries and archives to research institutes specializing in Holocaust research to data and information specialists, who are located in thirteen countries.¹⁶ Its target user group is researchers working in the area of Holocaust research.

Since they are sister projects, both designed to explore the challenges of providing support for research communities whose work is based on archival research, EHRI and CENDARI share certain features. One of those is a work package dedicated to the collection and analysis of user requirements. The EHRI project employs a wide variety of methods to collect information about

12 The authors would like to thank Petra Links of the NIOD Institute for War, Holocaust and Genocide Studies, Amsterdam, The Netherlands, for providing more information about the EHRI user requirement process and access to a pre-publication draft of their article (see Speck/Links *forthc.*). – See also the contributions to this volume by Blanke et al. (pp. 157 ff.), and Blanke/Fritze (pp. 243 ff.).

13 See the EHRI Factsheet, available at http://www.ehri-project.eu/drupal/web-fm_send/123.

14 See the Seventh Framework Programme website: http://cordis.europa.eu/fp7/home_en.html.

15 ARIADNE, the Advance Research Infrastructure for Archaeological Dataset Networking in Europe, will begin in 2013. The goal of this project is to provide a single point of access for researchers in the field of archeology who are seeking to integrate scattered datasets; see <http://www.cendari.eu/the-digital-humanities-landscape-in-europe-overview/>. EHRI, CENDARI, and ARIADNE are all based on the DARIAH research infrastructure framework and are deliberately designed to work together on similar goals, serving different user groups, as a test case to develop administrative best practices and methods for humanities VREs.

16 For more information about the EHRI project, including project partners, see <http://www.ehri-project.eu/partners>.

its user community and their requirements.¹⁷ To become familiar with the Holocaust research community, EHRI members created a bibliography and a stakeholder report. In addition to desk research into the research practices of this community, the use of partner sites and related subject-specific resources was also a topic of investigation. To gather first-hand knowledge of their target users' needs, EHRI researchers engaged in face-to-face interviews with Holocaust researchers and archivists as well as a web-based survey. In addition to this initial collection of knowledge about user needs, ongoing communication channels with the user community that will continue throughout the project duration include an online community hub, expert meetings, a fellowship program to sponsor researchers working with EHRI, summer schools, newsletters, an email listserv, and online training programs.¹⁸ The EHRI user requirements also build on the information on user requirements collected in DARIAH (cf., for example, Stavros et al. 2010: 5–7). However, DARIAH is separate from more individual Digital Humanities projects in scope, since its activities focus on the establishment of infrastructures rather than on the development of specific analytical and methodological tools (cf. Romary 2013: 17).

EHRI's collection of user requirements includes one source of information in the form of user statistics that was to be provided by archives and partner institutions, since these organizations would be most familiar with their users and their requirements (cf. Speck/Links *forthc.*: 6). Since the amount of information archives and partner institutions record and were able to provide about their users varied widely, however, the original plan to gather this information directly in the form of statistical data was supplemented by interviews with individual archivists. This process resulted in a wealth of information about the importance of archivists as mediators between users and the collections and holdings of archives, both online and offline, which must be taken into account in creating digital resources (cf. *ibid.*: 13–14).

Another change in the approach to understanding user requirements involved the decision to establish a Research Users Advisory Committee in order to have continued feedback and input from a pool of Holocaust re-

17 See Work Package 16: "UserRequirements" at <http://www.ehri-project.eu/drupal/partners-organisation>.

18 More information about the EHRI community hub and mailing lists are available at <http://www.ehri-project.eu/drupal/subscription-info>.

searchers on the resulting documents dealing with the collection of user requirements. This committee was created to provide a type of focus group for interviews and to supply direct comments and suggestions, such as by testing and disseminating the web survey, to ensure that the user requirements accurately reflect the needs of the Holocaust researcher community.

EHRI relies on researcher involvement and channels of communication on various levels including the Research User Advisory Committee, meetings, online networking, and information hubs. As with TextGrid, encouragement for the adoption of standards within the researcher community, and the emphasis on long-term sustainability of data and tools for working with the data, are seen as important factors in accommodating future research scenarios and approaches.

Another approach for EHRI to anticipating future user requirements was revealed in the process of carrying out the informational interviews with archivists. Since archivists serve as mediators between users and the archival holdings, use of Web 2.0 communicative media could be used to reinvent the relationship between the historian and the archivist so that present and future needs of researchers are matched with the knowledge and guidance of the archivists (cf. *ibid.*). In this way, the requirements of future researchers dealing with new methodological approaches would be grounded in a larger context of improved communication and information exchange, in which researchers seeking information would be more closely connected with archivists providing access to sources and information about them. As a platform to potentially support this interaction, EHRI's role would be to enable communication between both groups and would as a result have a direct channel to users and their research processes within the context of the larger research community – information that could serve as the basis for future technological development.

The Collaborative European Digital Archive Infrastructure (CENDARI)

CENDARI is developing a portal to support collaborative research in the fields of World War I and medieval history, and to aid in archival research in geographically-disparate collections. CENDARI is funded by the European Commission's Seventh Framework Programme, with a project duration of four years (2012–2016), and has fourteen partners, consisting of libraries and

information specialists, research consortiums, and research institutes from the disciplines of WW I history and medieval studies, who are located in seven countries.¹⁹ Its target user groups are historians and researchers working in the domains of medieval studies and WW I, in addition to archivists, librarians, and information specialists.

Similar to EHRI, CENDARI uses a number of techniques to gather information about its user communities in the form of desk research into research methodologies and processes in online and offline contexts, explorations of partner resources and existing tools and services, online surveys, face-to-face interviews, and expert seminars to bring researchers together with librarians and archivists and infrastructure specialists. There is ongoing project-internal contact with researchers in medieval studies and WW I who are employed by the project and who represent their subject areas as well as project-external contact, both in the form of connections to existing user communities and through outreach programs such as fellowship programs and summer schools. The External Expert Advisory Board, comparable to EHRI's Research Users Advisory Committee, provides feedback on work within the project. There is also a work package that deals with communication and information dissemination, including administering the surveys, blog posts, Twitter and Facebook feeds, and providing notifications of publications about CENDARI.

One additional method for the CENDARI project's collection of user requirements were the participatory design workshops held for each of the target user communities. Participatory design is a technique used in interactive software design to ensure that users and designers communicate effectively so that the final product fulfills the desired purpose (cf., for example, Edwards et al. 2007: 34–35). The point is to create a common point of reference for differing perspectives, since users are familiar with their research and its contexts, while developers know about the available technology. Participatory design sessions use techniques such as brainstorming and video prototyping to encourage participants to think innovatively and to map the interactions between people and computers. Based on the materials developed by Wendy McKay of INRIA (cf. McKay n.d.),²⁰ these workshops focus

19 For more information about CENDARI, including the project partners, see <http://www.cendari.eu/sample-page/about-us/>.

20 For video tutorials explaining the participatory design process using video prototyping, see <http://www.lri.fr/~mackay/VideoForDesign/>.

on the creation of specific prototypes and sample interfaces by means of paper and video cameras to create actual concrete scenarios of how a user interacts with an interface to carry out a specific task. The results of these sessions, both in the form of tangible video prototypes and mockups and the intangible exchanges between researchers and developers, are also currently feeding into the creation of user stories, which are short statements used in agile software development to describe user needs in a way that is clear to both end users and developers (cf., for example, Ramsay 2011).

In anticipating the needs of future users, both CENDARI and EHRI are working together as a type of case study to investigate archival research practices. In creating a VRE for different user communities that all depend on archival research, this is a way to explore and consider the administrative best practices and common functions that can be extended to broader contexts of research. In the long run, this combined approach will result in more information about which approaches are effective and best support archival research in the online and offline resources.

Conclusion

All of the projects overviewed here were planned with the intention of having an influence on future methods and research processes, making communication between researchers and developers a dialogue rather than a one-way channel. TextGrid, EHRI, and CENDARI all rely on the interaction between research communities, information specialists such as librarians and archivists, and technical specialists to provide an understanding of user requirements.

Whether it is possible to “program for the future” since user requirements change drastically, as do disciplinary methodologies, research topics and themes, and funding opportunities, is perhaps not as important as another question: what will happen if we do not try? Anticipating future needs is necessarily quixotic but nevertheless vital in project development. Just as communication between researchers and developers and information specialists must be an ongoing dialogue, concepts and planning for projects must take into account present needs and future potential possibilities. The tools used now and the data that is made available and maintained will affect the research questions and issues that can be raised in the future; projects who

ignore their role in this development risk programming neither for the future nor the present.

Literature

- Carusi, Annamaria; & Torsten Reimer (2010): Virtual Research Environment Collaborative Landscape Study: AJISC Funded Project. <http://www.jisc.ac.uk/publications/reports/2010/vrelandscapestudy.aspx>.
- Courage, Catherine; & Kathy Baxter (2005): *Understanding your users. A practical guide to user requirements*. San Francisco: Morgan Kaufman.
- Ewards, Paul N.; Jackson, Steven J.; Bowker, Geoffrey C.; & Cory P. Knobel (2007): Understanding Infrastructure: Dynamics, Tensions, and Design. Report of a Workshop on History & Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures. <http://deepblue.lib.umich.edu/bitstream/handle/2027.42/49353/UnderstandingInfrastructure2007.pdf?sequence=3>.
- Grady, Jeffrey O. (2006): *System Requirements Analysis*. Burlington, MA: Academic Press.
- Küster, Marc W.; Ludwig, Christoph; & Andreas Aschenbrenner (2007): TextGrid as a Digital Ecosystem. In: *2007 Inaugural IEEE Conference on Digital Ecosystems and Technologies (IEEE DEST 2007)*. pp. 506–511.
- Liess, Sabine; Aschenbrenner, Andreas; & Martina Kerzel (2006): Öffentlichkeitsarbeit R 6.1. <http://www.textgrid.de/ueber-textgrid/materialien/antraege-und-beichte/>.
- McKay, Wendy E. (n.d.): Using Video to Support Interactive Design. <https://www.lri.fr/~mackay/VideoForDesign/print/print.pdf>.
- Ramsay, Anders (2011): The UX of User Stories, Part I. <http://www.andersramsay.com/2011/07/16/the-ux-of-user-stories-part-1/>.
- Romary, L. (2013): Technologies, services and user expectations – Prospects for DARIAH (DARIAH Report EU 4.3.2; version from 15 February 2013). <http://hal.inria.fr/hal-00912653>.
- Schwiegelshohn, Uwe (2008): D-GRID: The communities of Germany's D-Grid. In: *Strategies Projects 7*, pp. 14–18.
- Speck, Reto; & Petra Links (forthcoming): The Missing Voice: Archivists and Infrastructures for Humanities Research (submitted to the *International Journal of Humanities and Arts Computing*).

Stavros, Angelis et al. (2010): DARIAH Technical Report – overview summary.
http://subugoe.github.io/website-DARIAH-EU/indexc22d.pdf?option=com_docman&task=doc_download&gid=477&Itemid=200.

Sutcliffe, Alistair (2002): *User-Centred Requirements Engineering*. London: Springer.

Theodor Fontanes Notizbücher

Genetisch-kritische und kommentierte Hybrid-Edition,
erstellt mit der Virtuellen Forschungsumgebung TextGrid

von Gabriele Radecke, Mathias Göbel und Sibylle Söring

1 Einleitung

Notizbuch-Editionen erfordern komplexe Methoden der Transkription, der Textkonstitution und der Kommentierung, die sich an der Materialität und Medialität sowie an den Funktionen und Inhalten der Überlieferungsträger orientieren müssen. Konventionelle Verfahren, die sich lediglich auf die Inhalte von Notizbuchaufzeichnungen konzentrieren und zum Ziel haben, die Notate wohlgeordnet in einer Buchedition zu veröffentlichen, haben nicht nur sehr begrenzte Möglichkeiten, die medialen, funktionalen und materialen Eigenschaften von Notizbüchern aufzubereiten. Sie verhindern auch, dass alle substanziellen Merkmale, die die Notizbücher in ihrer einmaligen physischen Gestalt charakterisieren und von anderen Handschriftenkonvoluten maßgeblich unterscheiden, von den Lesern und Benutzern rezipiert und für Forschungszwecke ausgewertet werden können.

Zwar wurden innerhalb der traditionellen neugermanistischen Editions-wissenschaft die Transkriptionsprinzipien seit D. E. Sattlers Frankfurter Hölderlin-Ausgabe immer weiter ausdifferenziert und grafisch umgesetzt (vgl. Hoffmann/Zils 2005: 221–227), aber erst die digitalen Methoden, die seit nunmehr zwei Jahrzehnten die Editionsphilologie bereichern, ermöglichen gleichermaßen eine annähernd überlieferungsadäquate zeichen- und positionsgetreue Darstellung, Auszeichnung und Codierung. Standards hierfür setzten Projekte wie HyperNietzsche, deren Herausgeber das Verfahren und den Begriff der „ultra-diplomatischen Transkription“ entwickelten, die Transkription der dritten Art:

“The third kind of transcription, the ‘ultra-diplomatic transcription’, requires some human intervention, as it represents very detailed physical features in an ‘iconic’ way, which can hardly be encoded by markup and displayed in HTML: the exact positioning of words, passages and graphics on the paper, the exact font size and the like.” (Saller 2003)

Auch die im Folgenden vorgestellte Edition der Notizbücher Fontanes orientiert sich an diesem anspruchsvollen Maßstab. Im Unterschied zum HyperNietzsche-Vorhaben, in dem die Transkriptionsansicht unter Zuhilfenahme einer browserintegrierten PDF-Ansicht umgesetzt werden sollte, werden die beiden ‚Text‘-Ansichten – die diplomatisch-dokumentarische Transkription und der zeichengetreu-linear edierte Text – sowie die Apparate aus einer gemeinsamen Datengrundlage heraus transformiert und im Browser angezeigt, ohne dass zusätzliche, texttechnologisch nicht erschlossene Dateiformate wie etwa PDFs erzeugt werden müssen. Die Idee ist freilich keine neue. Die Ursprünge finden sich bei Robert Busa, dessen „Index Thomisticus“ im Jahre 1949 als erste elektronische Edition gilt (vgl. Busa 1980; Jannidis 2008: 317 f.).¹ Durch die Zusammenarbeit mit IBM, dem industriellen Vorreiter des digitalen Zeitalters, wurde eine Brücke gebaut, die auch heute noch zwei so verschiedene Wissenschaften miteinander verbindet, vor der aber auch der technische Fortschritt nicht haltgemacht hat. Was sich seit der Verwendung von Lochkarten bis zur heutigen Architektur webbasierter Editionen verändert hat, soll en passant auch aus diesem Beitrag deutlich werden.

2 Voraussetzungen

2.1 Literatur- und editionswissenschaftliche Voraussetzungen

Theodor Fontanes Notizbücher sind das letzte, noch unveröffentlichte größte Textkorpus des Autors. Obwohl 67 Notizbücher überliefert sind, die Fontane über Jahrzehnte hinweg geführt hat, sind diese nicht als *echte* Notizbücher rezipiert worden. Verantwortlich dafür ist die desolante Editionssituation, denn bisher wurden nur wenige Aufzeichnungen aus dem Zusammenhang herausgenommen und in unterschiedlichen Buchausgaben publiziert. So gibt es nur einige, zumeist populäre Teilpublikationen,² und die anderen, bisher unveröffentlicht gebliebenen Notizbuchinhalte sind auch nur

1 Dies gilt auch für die Historisch-Kritische Kafka-Ausgabe, die unter dem „elektronischen Teil der Edition [...] weitgehend [...] eine möglichst getreue Wiedergabe des gedruckten Textes“ versteht. Auch hier hat man das PDF-Format verwendet, wobei die Texte allerdings auf CD-ROM abgespeichert wurden; vgl. Jannidis (2005: 467).

2 Vgl. die Zusammenstellung in Radecke (2008: 211–214) und Radecke (2010: 95–97).

gelegentlich von der editions- und literaturwissenschaftlichen Forschung genutzt worden. Selbst die verdienstvollen Fontane-Standardwerke, beispielsweise das „Fontane-Handbuch“³ und die „Fontane-Chronik“ (Berbig 2010), haben auf eine Auswertung aller Notizbuchaufzeichnungen verzichtet. Der von der Editionswissenschaft inzwischen erarbeitete Standard für Editionen literarischer Texte, die eine systematische Sichtung der handschriftlichen Textzeugen, die Transkription, die genetische Dokumentation und Textkonstitution umfassen, ist innerhalb der Fontane-Forschung nicht einmal annähernd erreicht worden.⁴ Obwohl Fontanes poetisches und journalistisches Werk in einer dichten und disparaten handschriftlichen Überlieferung vorliegt, gibt es zwar vier Studienausgaben, aber keine Ausgabe, die die Handschriften – und somit auch die Notizbuchaufzeichnungen – historisch-kritisch oder genetisch präsentiert und der Materialität des Überlieferungsträgers angemessen Rechnung trägt.

Hinzu kommt, dass die wenigen Auszüge von Notizbuchniederschriften in den Fontane-Werkausgaben nicht innerhalb einer Rubrik „Notizbücher“, sondern auf verschiedene Abteilungen zerstreut veröffentlicht und somit in neue Kontexte eingebunden wurden. Diese Entscheidung hat maßgeblich dazu beigetragen, dass Fontanes Notizbücher nicht als zu Fontanes Werk gehörig rezipiert worden sind und keinen Eingang in das kulturelle Gedächtnis gefunden haben. In der „Nymphenburger Ausgabe“ sind einige der Notizbuchaufzeichnungen im Kommentarteil versteckt abgedruckt.⁵ Die „Hanser-Ausgabe“ bietet unter dem missverständlichen Titel „Tagebücher“ Kostproben der in den Notizbüchern enthaltenen tagebuchartigen Reiseaufzeichnungen; sie stellt also ohne editorische Transparenz Fontanes Notizbuchauf-

3 So fehlt etwa im „Fontane-Handbuch“ ein Kapitel zu den Notizbüchern. Gotthard Erler weist lediglich in einem kurzen Abschnitt auf Fontanes sogenannte Reisetagebücher hin; vgl. Erler (2000: 771 f.). Dieser Begriff wurde von Friedrich Fontane erstmals verwendet und hat sich auch innerhalb der Forschung durchgesetzt, obwohl er nicht zutrifft. Denn die heterogenen tagebuchartigen und weiteren Aufzeichnungen (Exzerpte, Zusammenfassungen von Reise-Eindrücken und Alltagsnotizen), die während der Unternehmungen entstanden sind, schrieb Fontane nicht in seine Tagebücher, sondern in die Notizbücher, die Fontane selbst stets als „Reisenotizbücher“ bezeichnete.

4 Zu den vielschichtigen Ursachen vgl. Radecke (2002: 151–155) und zuletzt Radecke (2010: 95 f.).

5 Vgl. NFA, Bd. XVIIIa (1972), Reisebriefe aus Jütland, S. 917–931, Die Rheinreise, S. 1135–1161; NFA, Bd. XXIII/2 (1970), Reisen nach Italien 1874/75.

zeichnungen neben die in seinen Tagebüchern festgehaltenen Niederschriften.⁶ Auch innerhalb der noch nicht abgeschlossenen „Großen Brandenburger Ausgabe“ wurden einzelne Notizbuchniederschriften lediglich auf die Abteilungen der „Gedichte“, der „Wanderungen“ und der „Tage- und Reise-tagebücher“ verteilt abgedruckt. Schließlich nutzte man die Notizbücher gelegentlich für die Kommentierung, etwa der „Wanderungen durch die Mark Brandenburg“, der „Gedichte“ und des erzählerischen Werkes.⁷

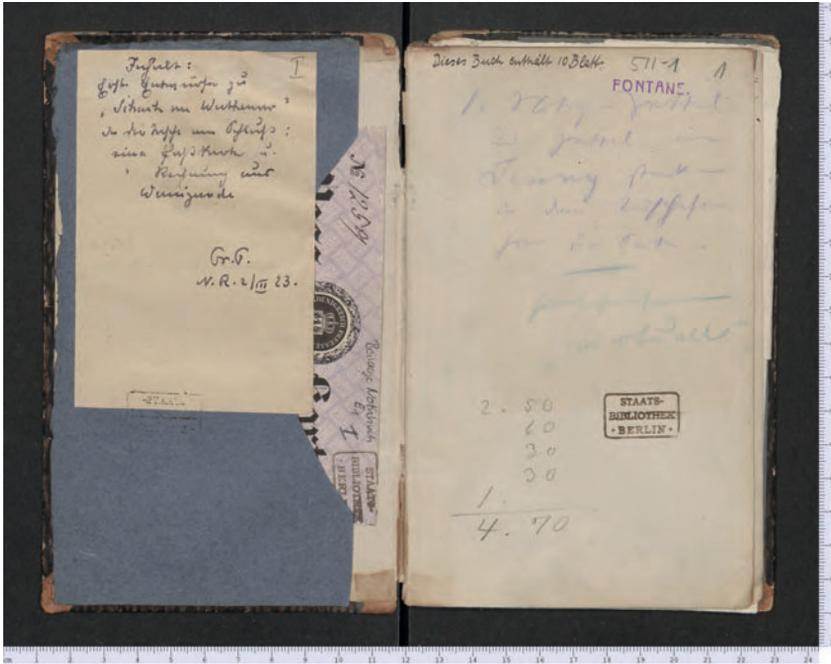


Abb. 1 Notizbuch E01, vordere Einbanddecke v: Inhaltsverzeichnis, von Friedrich Fontane angelegt

6 Vgl. HFA, Abt. III, Bd. 3/II (1997). Dort wurden Fontanes Notizbuchaufzeichnungen aus Dänemark (1864), der Rheinreise (1865), der Thüringenreisen (1867/1873) und der Italienreisen (1874/75) erneut veröffentlicht.

7 Vgl. AFA – Gedichte (1989), AFA – Romane und Erzählungen (1973) und AFA – Wanderungen (1976–1987) sowie GBA – Das erzählerische Werk (1997 ff.), GBA – Gedichte (1995), GBA – Tage- und Reisetagebücher (1994–2012) sowie GBA – Wanderungen (1997).

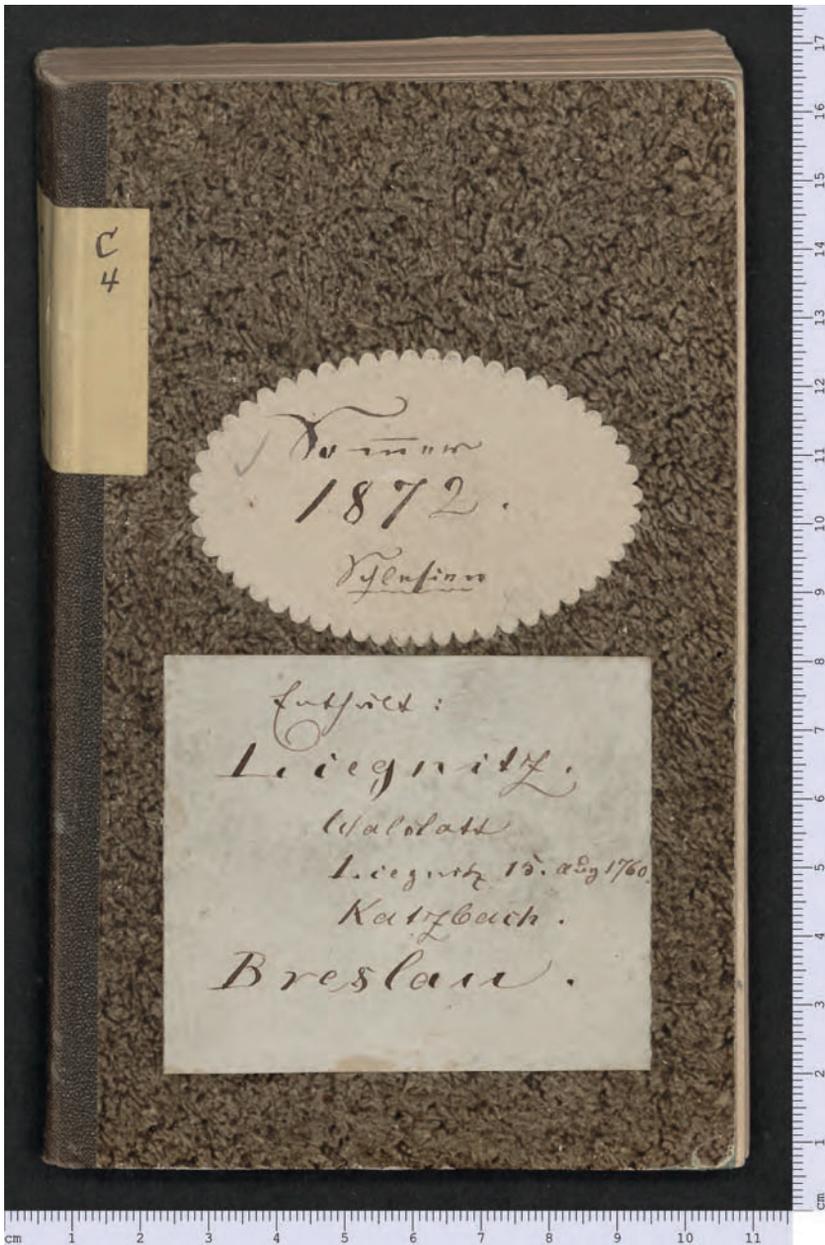


Abb. 2 Notizbuch C04, vordere Einbanddecke r: Sommer 1872 Schlesien. Fontanes Inhaltsübersicht

Um 1923 hat Friedrich Fontane, der jüngste Sohn Theodor Fontanes, eine erste Bestandsaufnahme der Notizbücher auf den vorderen inneren Einbanddecken vorgenommen (vgl. Abb. 1). Er ergänzte die von seinem Vater auf den äußeren Einbanddecken angelegten, aber längst nicht alle Aufzeichnungen erfassenden Themen- und Inhaltsübersichten (vgl. Abb. 2). 1976 wurde in den „Fontane Blättern“ eine Übersicht veröffentlicht, der allerdings keine Autopsie vorausging, sondern die lediglich die einzelnen, von Theodor und Friedrich Fontane angelegten fehler- und lückenhaften Informationen zusammenführte (vgl. Schubarth-Engelschall 1976).⁸

Obwohl also spätestens mit dieser Auflistung Hinweise auf die Inhalte der Notizbücher gegeben wurden, fehlt 115 Jahre nach Fontanes Tod immer noch eine Notizbuch-Gesamtedition, die auch eine systematische Erschließung der Inhalte umfasst. Es liegen hingegen nur wenige gedruckte, aus dem Überlieferungskontext herausgelöste Teilpublikationen von Notizbuchaufzeichnungen vor, die – allein durch inhaltlich-systematische Kriterien motiviert – in neue Text- und Werkzusammenhänge integriert worden sind. Das Medium Notizbuch mit seinen materialen Eigenschaften – der Papierqualität, dem Format, dem Nach- und Nebeneinander von beschrifteten und unbeschrifteten Seiten, den mehrschichtig beschrifteten Seiten (vgl. Abb. 3), den Schreibgeräten und -stoffen (vgl. Abb. 4), dem Duktus (vgl. Abb. 5) sowie den vakaten oder beschrifteten Blattfragmenten, den Beilagen (vgl. Abb. 6), den eingeklebten Zeitungsausschnitten (vgl. Abb. 7) und anderen Druckerzeugnissen sowie der Relation von Schrift und Bild (vgl. Abb. 8) – spielte für bisherige Editionsconzepte keine Rolle. Infolgedessen wurden alle Informationen, die durch die Analyse der Materialität der physischen Notizbücher hätten gewonnen werden können – etwa Fragen bezüglich der Textsorten und Gattungen (vgl. Radecke 2008; 2010) sowie der Datierung, der Schreibchronologie und der dichterischen Arbeitsweise – durch die bisherigen Drucke verwischt.⁹

8 Mit der Hybrid-Edition von Theodor Fontanes Notizbüchern wird erstmals ein autopsiertes Gesamtinhaltsverzeichnis erarbeitet.

9 Hinzu kommt, dass die Transkriptionsrichtlinien den Prinzipien derjenigen Editionen folgen, deren edierte Texte aufgrund gedruckter Textvorlagen konstituiert wurden. Die Notizbuchaufzeichnungen wurden infolgedessen nicht dem editionswissenschaftlichen Standard gemäß zeichengetreu mit allen Überarbeitungs- und Tilgungsspuren ediert, sondern zumeist in modernisierter, normierter und „endgültiger“ Gestalt mit zahlreichen Texteingriffen der Herausgeber.

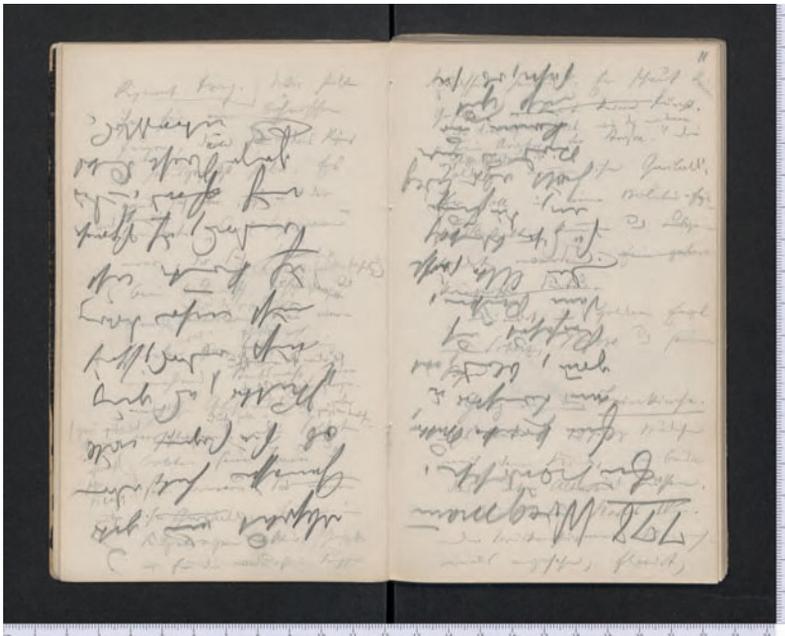


Abb. 3 Notizbuch D04, Blatt 10v/11r: doppelt beschriftete Aufzeichnungen



Abb. 4 Beispiele zu den Schreibgeräten/-stoffen

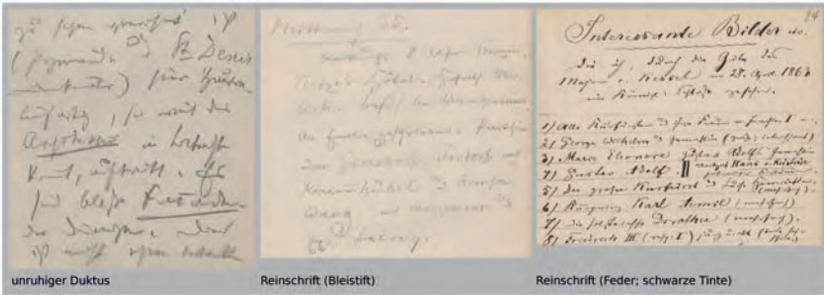


Abb. 5 Beispiele für unterschiedlichen Duktus



Abb. 6 Notizbuch E01, Beilage I: Passkarte 1874

Wie viele seiner Schriftstellerkollegen pflegte Fontane über einen längeren Zeitraum seines Schreibens Notizbücher zu führen. Von etwa 1860 bis Ende der 1880er-Jahre füllte er mindestens 67 Hefte, die zwischen 44 und 176 Blatt umfassen, unterschiedliche Formate haben und aus disparaten Materialien bestehen. Fontane hat diese Form der Gedächtnisstütze und Stoffsammlung bei mehreren Gelegenheiten eingesetzt: Er führte die Notizbücher auf seinen Reisen in die Mark Brandenburg, nach Schlesien, Italien und in

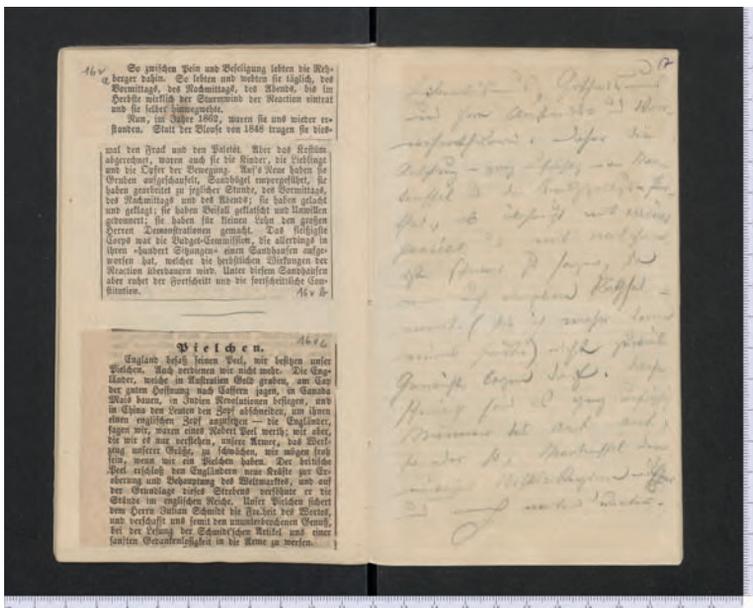


Abb. 7 Notizbuch A12, Blatt 16v: aufgeklebte Zeitungsausschnitte

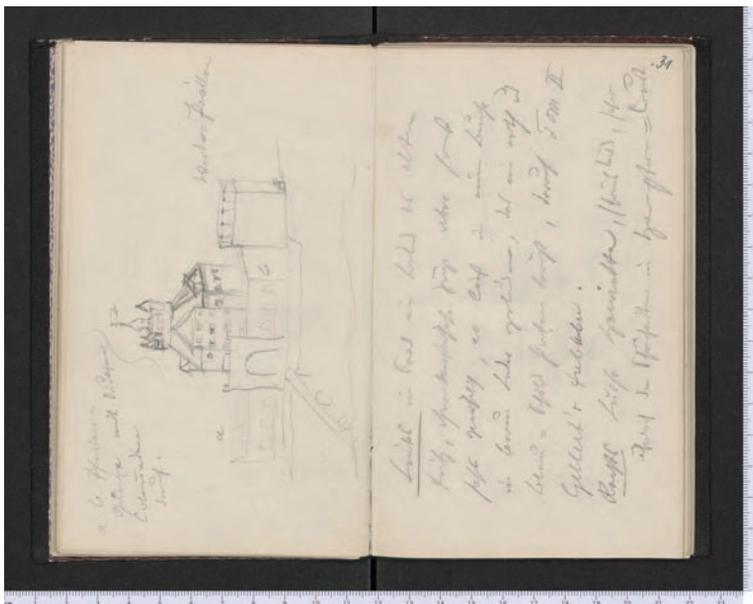


Abb. 8 Notizbuch A03, Blatt 30v/31r: Schloss Oranienburg

die Schweiz sowie bei den Erkundungen von Schlachtfeldern in Dänemark, Böhmen und Frankreich mit sich. Er hielt darin Eindrücke von Kunstausstellungen und Theateraufführungen fest, schrieb während der Lektüre von Büchern und bei Vorträgen mit oder konzipierte eigene Texte. Außerdem sind Briefkonzepte und To-do-Listen sowie Übersichten über Zug-Abfahrtszeiten und Honorare überliefert, die Einblicke in den beruflichen und privaten Alltag geben. Aus diesen medialen Bedingungen ergeben sich heterogene Aufzeichnungen: Neben Tagebucheinträgen finden sich Entwürfe zu Prosatexten und Gedichten; Exzerpte wechseln sich ab mit Reisebeschreibungen und Zeichnungen von Gebäuden, Grabdenkmälern, Stadtplanausschnitten und Umgebungskarten. Viele seiner Aufzeichnungen hat sich Fontane später für verschiedene Schreibprojekte erneut vorgenommen und überarbeitet (vgl. Abb. 9).

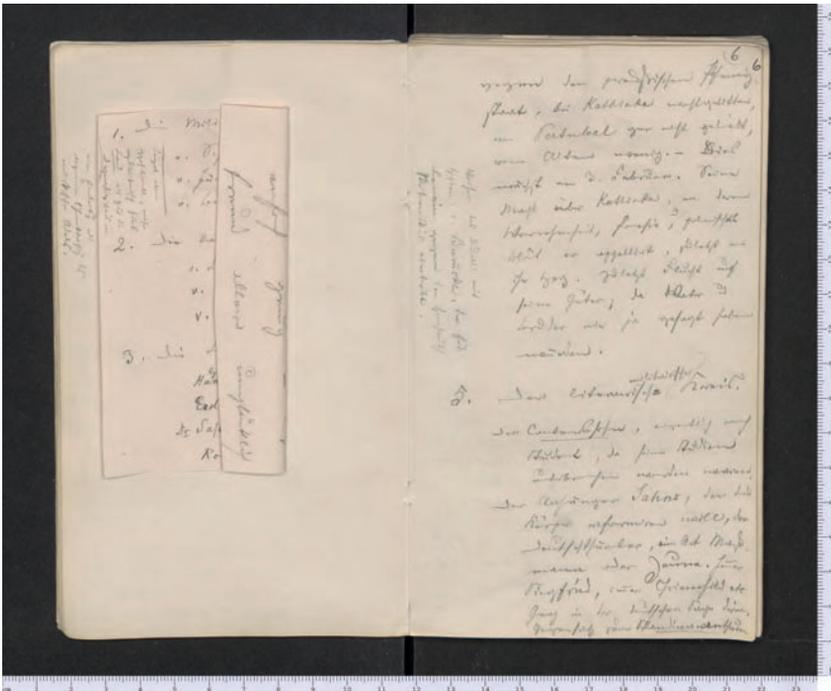


Abb. 9 Notizbuch E03, Blatt 6r: Beispiel für Überarbeitungsspuren

2.2 Informationswissenschaftliche Voraussetzungen

Um diese Heterogenität des Materials, der Inhalte und der Arbeitsweise innerhalb des Mediums Notizbuch nicht nur zu erfassen, sondern sie vielmehr auch adäquat abbilden und auswerten zu können, sind in der Informationswissenschaft viele Grundsteine gelegt. Es ist ein Fundament vorhanden, von dem nicht nur die tradiert datenintensiven Naturwissenschaften, sondern auch die bild- und textbasierten Geisteswissenschaften profitieren können; dazu zählen Technologien von der Herstellung qualitativ hochwertiger Digitalisate bis zur sicheren Langzeitarchivierung dieser Abbilder und ihrer zugehörigen Forschungsdaten. Zum interdisziplinären Austausch tragen Standards bei, die in den grundlegenden technischen Entwicklungen des Internets ihre gemeinsame Wurzel haben.

Die eXtensible Markup Language (XML) ist ein Beispiel hierfür. Auf der Grundlage dieses Datenformats führt die Text Encoding Initiative (TEI) auf ganz spezifische Weise die Informationswissenschaft mit der Geisteswissenschaft zusammen. Seit 1986 erarbeitet die TEI eine Systematik zur texttechnologischen Erschließung von materiell vorhandenen Schriftstücken, gleich welchen Formates, gleich welcher Art und gleich welchen Inhaltes; sie ist aber auch anwendbar auf Objekte, denen das Attribut ‚born digital‘ zugeschrieben werden kann. Die Empfehlungen des Konsortiums werden inzwischen weltweit von Editionsprojekten aufgegriffen und haben sich mittlerweile zu einem De-facto-Standard entwickelt (vgl. Jannidis 2008: 320).¹⁰ Auch die Auszeichnung der Metadaten, die digitale Abbildung der Transkriptionen und die Generierung verschiedener edierter Texte und Apparate sind mittels der TEI-Guidelines – seit 2012 in der 5. Version (P5) verfügbar – möglich. Von besonderer Relevanz für die in der Fontane-Edition vorliegenden Materiallage ist im Rahmen der TEI die Special Interest Group für Manuskripte, die die ständige Anpassung von Codierungsvorschlägen bzw. Guidelines für handschriftliche Zeichenfolgen zum Gegenstand hat.¹¹ Die daraus entstehenden projektspezifischen Document Type Definitions (DTD), die für die TEI innerhalb sogenannter ODD-Dokumente personalisierte Auszeichnungsschemata festlegen, haben den Umgang mit diesen Guidelines für Anwender stark vereinfacht. So sind nur solche Elemente und Attribute in-

10 Vgl. auch die Auflistung ausgewählter digitaler Editionsprojekte basierend auf XML und TEI auf der Website http://en.wikipedia.org/wiki/Text_Encoding_Initiative.

11 Vgl. <http://www.tei-c.org/Activities/SIG/Manuscript/>.

nerhalb eines Dokumentes gültig, welche auch durch eben dieses Schema beschrieben werden. In diesem Sinne handelt es sich um Einschränkungen der Metasprache. Seither nutzt man nicht mehr eine generalisierte DTD, sondern individuelle projektbezogene Auszüge, die bei Bedarf auch angepasst werden können.

Einen besonderen Vorteil von XML-Anwendungen bieten die daran anknüpfenden und ebenfalls vom World Wide Web Consortium (W3C) verabschiedeten Standards. Dazu zählen Transformationen in beliebige andere Formate mittels Extensible Stylesheet Language Transformation (XSLT), aber auch weitere turing-vollständige Programmiersprachen wie XQuery. Mit diesen angeschlossenen Programmiersprachen ist es möglich, basierend auf einer Datengrundlage einen flexiblen Output zu erzeugen, der sich je nach Anfrage an die prozessierende Instanz anpasst. Wissenschaftlerinnen und Wissenschaftler haben heute nicht nur eine globale Zugriffsmöglichkeit auf die Daten; sie können diese vielmehr auch prozessieren und je nach Forschungsschwerpunkt differenzierte Darstellungen generieren – unabdingbar bei der beschriebenen Materialheterogenität.

Damit liefert XML/TEI das Standard-Rahmenwerk zur Textauszeichnung und -codierung, wie sie insbesondere im Rahmen digitaler Editionen zur Anwendung kommt. Auf Seiten des Softwareangebotes fehlt es allerdings noch weitgehend an einer möglichst einfachen Umsetzung zur Erstellung solcher Editionen. Eine der Voraussetzungen dafür, Forscherinnen und Forschern das Potenzial solcher digitalen Unterstützung geisteswissenschaftlicher Arbeitsprozesse und Methoden zu eröffnen, ist etwa eine möglichst überschaubare, intuitive grafische Oberfläche, die einen schnellen Einstieg ermöglicht und bisherige Arbeitsabläufe wesentlich erleichtert. Der Einstieg in den Umgang mit diesen Technologien ist in der Regel zu begleiten; gerade in der Anfangsphase ist darüber hinaus technischer Support erforderlich. Diese Lücke schließt die Virtuelle Forschungsumgebung „TextGrid“, die explizit den Workflow zur Erstellung einer digitalen Edition bedient: von der Erfassung des digitalen Materials, dem nach Open-Archive-Information-System-Terminologie (OAIS) sogenannten Ingest über die Codierung und weitere elektronische Bearbeitung – wie Indexierung und Lemmatisierung – bis zur elektronischen Publikation in externen Portalen, in Datenbanken und im TextGrid-eigenen digitalen Langzeit-Archiv (TextGrid Repository). Die Ausgabe eines Dokuments zum Druck wissenschaftlicher Texte auch mit komplexen satztechnischen Layoutanforderungen ist ebenfalls möglich.

Mit O AIS ist eine Instanz angesprochen, die für den hier nur kurz skizzierten Workflow von besonderer Bedeutung ist. Als Referenzmodell digitaler Archivierung liefert O AIS nicht nur einen technischen Standard, es benennt und definiert auch die Begrifflichkeiten und schafft somit eine weltweit akzeptierte computersprachliche Grundlage. Auf ähnlicher Basis wäre auch eine gemeinsame Sprache von Geistes- und Informationswissenschaftlern zu entwickeln: “[B]ringing arts and humanities researchers together with information technology specialists in order to develop a common language and culture” (Neuroth/Lohmeier/Smith 2011: 230). Wo Geisteswissenschaftler neue Technologien nutzen, die für bisherige Arbeitsprozesse und -methoden einen dezidierten Mehrwert darstellen, und Informationswissenschaftler projektspezifische und auf ‚usability‘ ausgelegte Produkte entwickeln, die philologische Methoden, Prinzipien und Arbeitsprozesse abbilden, sind die besten Voraussetzungen für eine gute und weiterführende interdisziplinäre Zusammenarbeit gegeben, die das Potenzial hat, u.a. auch neue Forschungsfragen zu stellen und zu beantworten.

3 Rahmenbedingungen des Forschungsprojekts

Das Fontane-Editionsprojekt wird von der Deutschen Forschungsgemeinschaft gefördert und entsteht unter der Leitung von Gabriele Radecke und Heike Neuroth an der Theodor Fontane-Arbeitsstelle der Universität Göttingen sowie der SUB Göttingen, Abteilung Forschung & Entwicklung.¹² Die erste Projektphase (2011–2014) hat die Erarbeitung der Editionsprinzipien, die Konzeption des Metadatenschemas, die Vorbereitung und Durchführung der Transkription, der Textauszeichnung und der Codierung sowie die Konzeption des elektronischen Portals zum Gegenstand. Erste Zwischenergeb-

12 Weitere Projektmitarbeiter sind: Mathias Göbel (IT/Visualisierung), Martin de la Iglesia (Metadaten), Judith Michaelis (editorische Assistenz) und Rahel Rami (editorische Assistenz). Beratung: Alexander Jahnke (Metadaten) und Sibylle Söring (IT/Visualisierung). Vgl. die Projektwebsite der Theodor Fontane-Arbeitsstelle <http://www.uni-goettingen.de/de/303691.html> sowie die Seiten der SUB Göttingen <http://www.sub.uni-goettingen.de/projekte-forschung/projektetails/projekt/theodor-fontane-notizbuecher/> und der DFG <http://gepris.dfg.de/gepris/OCTOPUS/?jsessionid=939E762DDBDEFDFEBFF973533A610DE9?context=projekt&id=193395716&module=gepris&task=showDetail/>. Die Staatsbibliothek zu Berlin, Preußischer Kulturbesitz, ist assoziierter Partner.

nisse wurden auf dem TextGrid Summit am 15. Mai 2012 vorgestellt.¹³ In einer zweiten Projektphase (voraussichtlich 2014–2017) werden die Prinzipien der Kommentierung für die digitale Edition und die Buchausgabe erarbeitet und umgesetzt (Überblickskommentare, Stellenkommentar, Register). Im Unterschied zu allen bisherigen, ausschließlich an inhaltlichen Kriterien orientierten Buchpublikationen ausgewählter Notizbuchaufzeichnungen stellt das Konzept der genetisch-kritischen Hybrid-Edition die komplexe Überlieferung mit ihren materialen und medialen Kennzeichen in den Mittelpunkt. Die Edition besteht aus zwei Teilen, die in abgestufter Weise die Materialität und Medialität visualisieren und dokumentenorientierte, chronologische und werkbasiertere Zugriffe ermöglichen sowie einen linearen, les- und zitierbaren Text herstellen: Die digitale Edition wird alle Notizbuchaufzeichnungen in synoptischer Darstellung von Digitalisat und diplomatischer Transkription sowie einen zeichengetreuen historisch-kritischen Text mit textkritischem Apparat, Überblicks- und Stellenkommentaren bereitstellen; die kommentierte Buch-Edition wird die historisch-kritische Textfassung mit Apparat, Kommentaren und ausgewählten Faksimiles enthalten.

Im Zentrum der Editionsarbeit steht die digitale Repräsentation der Forschungsergebnisse, die elektronische Publikation. Die synoptische Darstellung von Digitalisat, diplomatischer Transkription und ediertem Text sowie die historischen Überblicks- und Stellenkommentare sind zwei der wichtigsten Ziele der Ausgabe. Außerdem sollen eine Ansicht der Schreibchronologie und die Textgenese aufgrund von Apparaten und Animationen aufgerufen werden können. Verschiedene Suchoptionen (Volltextsuche, Codierungen, Personen und ihre Werke, Orte und Sehenswürdigkeiten sowie Institutionen) werden die Notizbuchinhalte auf verschiedenen Ebenen erschließen. Die durch die Apparate und Kommentare gegebenen Informationen werden durch Tooltips und eine Randspalte realisiert. Hierzu werden für die Portalentwicklung neben den schon genannten XML-Technologien auch HTML5 und CSS3 eingesetzt. Ein Geobrowser wird die Visualisierung der Geodaten unterstützen; als Datengrundlage werden sowohl die in den Notizbüchern genannten, durch den philologischen Kommentar verifizierten Orte als auch die bekannten Reiserouten Fontanes und schließlich die Entstehungsorte der Notizbuchaufzeichnungen dienen. Die generell dem Notizbuchmedium geschuldete Entitätendichte kann insbesondere bei Geoinformationen durch die

13 <http://textgrid.de/fileadmin/presentationen/tg-summit-2012/presentation-fontane.pdf/>

Nutzung offener Ressourcen oder auch dem Text zu entnehmenden Zeitinformationen mittels geografischer und gleichzeitig temporaler Visualisierung, wie sie der DARIAH-Geo-Browser¹⁴ bereitstellt, aggregiert werden.

Die Nachnutzung der Forschungsdaten unter der Bewahrung der „geistigen Urheber der geleisteten Arbeit“ (Kamzelak 2012: 209) bildet einen weiteren Aspekt der elektronischen Erschließung. Als Open-Access-Publikation ist der Zugang zum – bislang unerschlossenen – Korpus nicht länger regional oder national begrenzt. Es ist ein weiterer Schritt, der Forschung ortsunabhängig machen kann, wenn bestimmte, leider noch wenig evaluierte Qualitätsstandards, erreicht und eingehalten werden.¹⁵ Dementsprechend sollen nicht nur die Bild- und XML-Daten frei verfügbar, sondern auch offene Schnittstellen angeboten werden. Sie ermöglichen das Anknüpfen späterer Projekte an nach philologischen Qualitätsstandards erschlossene Materialien, die eine gute Dokumentation vorweisen können.

4 Projektspezifischer Workflow

Eine Virtuelle Forschungsumgebung wie TextGrid – in Deutschland bislang Vorreiter auf diesem Gebiet – ist dazu geeignet, den gesamten, hier skizzierten Forschungsprozess zu unterstützen. Sie stellt die Infrastruktur bereit, sodass die editionsphilologische Arbeit den zentralen Part ausmachen und mit den informationswissenschaftlichen Teilbereichen innerhalb einer sicheren und abgeschlossenen Umgebung verbunden werden kann.

Abbildung 10 stellt diesen Workflow für das Projekt „Genetisch-kritische und kommentierte Hybrid-Edition von Theodor Fontanes Notizbüchern basierend auf einer Virtuellen Forschungsumgebung“ dar. Der Workflow beinhaltet drei wesentliche Arbeitsrahmen: die Bereitstellung und Erschließung des Materials, die Aufbereitung des Materials in der Virtuellen Forschungsumgebung und die verschiedenen Nutzungsmöglichkeiten in digitalen und analogen Publikationen.

14 Der DARIAH-Geo-Browser ist erreichbar unter <http://dev2.dariah.eu/e4d>.

15 so beispielsweise die „Qualitätskriterien für elektronische Quellenedition“ von Porta Historica: <http://www.portahistorica.eu/editions/qualitaetskriterien-fur-elektronische-quelleneditionen>, oder auch die von Roland S. Kamzelak im Auftrag der „Arbeitsgemeinschaft für germanistische Edition“ zusammengestellten „Empfehlungen zum Umgang mit Editionen im digitalen Zeitalter“ (Kamzelak 2012) und http://www.ag-edition.org/empfehlungen_editionen_v01.pdf/

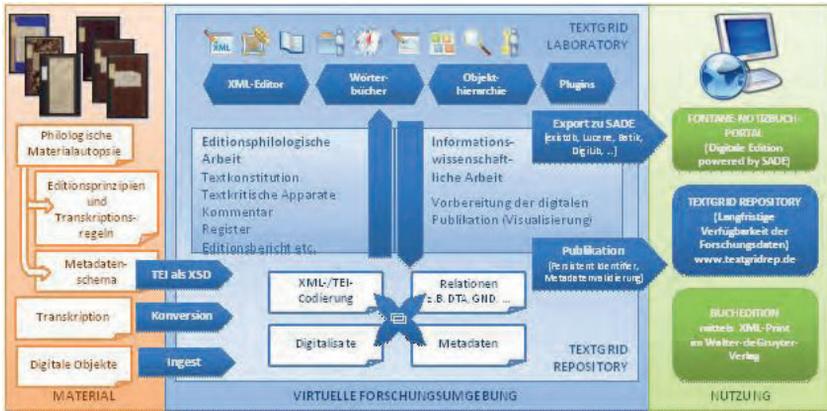


Abb. 10 Angepasster und erweiterter TextGrid-Workflow; Grafik erstellt vom TextGrid-Team, ergänzt und bearbeitet von Gabriele Radecke, Martin de la Iglesia und Mathias Göbel

Bereits die Materialautopsie und -bereitstellung erfordern eine enge Zusammenarbeit beider Disziplinen: Die philologische Materialautopsie bildet die Grundlage für die Editionsprinzipien, für die Transkriptionsregeln und für das Metadaten-schema. In dem sich anschließenden Arbeitsprozess ergänzen sich die Transkriptionsregeln und das Metadaten-schema gegenseitig, womit die editionsphilologischen, theoretischen und technischen Grundlagen geschaffen werden. Auch die Erstellung¹⁶ und die Qualitätskontrolle der Digitalisate erfolgt in interdisziplinärer Zusammenarbeit. Metrische Referenzen auf den Scans ermöglichen eine genaue bzw. relative Positionierung der Blattbeschriftung. Diese qualitativen Anforderungen bilden die Grundlage für die anschließenden Transkriptions- und Codierungsarbeiten, bei der einige XML-Elemente mit millimetergenauen Angaben versehen werden.

¹⁶ Da die Erstdigitalisierung durch das Theodor-Fontane-Archiv lediglich aus konservatorischen Gründen angefertigt wurde, bei der die fehlerhafte Foliierung übernommen wurde und der Großteil der unbeschrifteten Blätter und der Blattfragmente unberücksichtigt geblieben ist, wurde eine Neufolierung und Neudigitalisierung unter der Leitung von Gabriele Radecke und Jutta Weber von der Staatsbibliothek zu Berlin, Preussischer Kulturbesitz, veranlasst. Erst mit der Neudigitalisierung sind die Notizbuchaufzeichnungen auch im digitalisierten Zustand für den Benutzer als Teile des Mediums Notizbuch erfassbar und insofern für eine komplexe, materialbasierte Notizbuch-Edition brauchbar.

Mit dem Einspielen der Daten (Ingest) in das TextGrid Laboratory beginnt die eigentliche Arbeit in der Virtuellen Forschungsumgebung. Dazu zählen sowohl die Digitalisate als auch das projektspezifische Metadaten-schema, das im W3C-Schema-Format von XML-Editoren validiert und während der Codierungsarbeit weiter angepasst werden kann. Innerhalb des TextGridLabs ist die Bedeutung des XML-Editors als Werkzeug zur Codierung der Notizbuchaufzeichnungen hervorzuheben. Mittels der TextGrid-URI erfolgt, gemäß XSLT-Standard (vgl. Kay 2007), ein Verweis auf eine Transformationsdatei. Dieses Schema wird durch den innerhalb der TextGrid-Umgebung verfügbaren XSLT-Prozessor angewendet und generiert eine Vorschauansicht in Form einer Website. Dabei stellt die Verwendung von XSLT 2.0 eine wichtige Verbesserung dar, die direkt aus dem Fontane-Editionsprojekt an die Entwickler herangetragen wurde.

Die Flexibilität der Virtuellen Forschungsumgebung zeigt sich nicht nur im Hinblick auf die allgemeinen Dienste; auch projektbezogene Aktualisierungen sind permanent möglich. Nach dem Ingest können nunmehr alle erforderlichen Änderungen direkt im TextGridLab vorgenommen werden, was die Transparenz und die interdisziplinäre Zusammenarbeit fördert: Das Transkriptionsschema ebenso wie die Transkriptionsrichtlinien des Editions-vorhabens erfordern kontinuierliche Überprüfung, Aktualisierung und Anpassung. Das Metadaten-schema als eine die einzelnen XML-Dokumente übergreifende Instanz bleibt, jeweils in der zuletzt gültigen Version, mit diesen assoziiert. Der Übergang von der philologischen Erarbeitung der Transkription und Textauszeichnung sowie der Codierung zur technischen Realisation zeichnet sich somit durch einen hohen Grad an Transparenz, Sicherheit und Komfort aus: Projektmitarbeiterinnen und -mitarbeiter können jeweils automatisch auf die jüngste Version eines Dokuments bzw. einer Datei zugreifen und diese weiterverarbeiten. Das gilt im derzeitigen Projektstatus insbesondere für die Generierung der diplomatischen Transkriptionsansicht, für die – wie auch für die Erstellung des edierten Textes – XSLT-Dateien eingesetzt werden. Parallel zur Transkription und Codierung kann die Entwicklung der Transformationsszenarien laufen. So ist es möglich, manuelle Kollationen der Transkriptions- und Codierungsergebnisse mittels der transformierten und lesefreundlichen diplomatischen Oberflächenansicht im Browser bereits zu Beginn der Arbeit im TextGridLab auszuführen. Dazu werden zusätzlich relevante Elemente mittels JavaScript und per Buttons in der Oberflächenansicht hervorgehoben. So lassen sich die Eingaben schnell erfassen und überprüfen. Transkriptions- und Codierungsfehler können auf

diese Weise früh erkannt und weitgehend beseitigt werden. Testmaterial ermöglicht zudem eine kontinuierliche Erweiterung und Verbesserung des XSLT-Stylesheets, das wiederum ohne Zeitverzögerung für die Codierung der Notizbuchaufzeichnungen zur Verfügung steht. Die Arbeit mit revidierten Dateien gewährleistet die Nachvollziehbarkeit und Dokumentation solcher Änderungen und Erweiterungen.

5 Fazit

Die genetisch-kritische Hybrid-Edition der Notizbücher Fontanes, einem der bedeutendsten deutschsprachigen Autoren des 19. Jahrhunderts, wird erstmals Teile seines Werkes digital erschließen und veröffentlichen. Insbesondere durch die interdisziplinäre Zusammenarbeit von Editions- und Informationswissenschaft sowie der Informatik werden neue und entscheidende Impulse für die werkgenetische, literatur- und kulturwissenschaftliche sowie für die mentalitätsgeschichtliche Forschung erwartet. Denn erst gemeinsam entwickelte philologische und informationstechnologische Methoden ermöglichen Publikation, Analyse und Rezeption der Aufzeichnungen entsprechend ihrer überlieferten Materialität im Medium Notizbuch. Darüber hinaus wird die digitale Edition sowohl in ihren editionsphilologischen Rahmenbedingungen als auch in informationstechnologischer Hinsicht als Modell für zukünftige Notizbuch-Editionen anderer Autoren mit ähnlich schwierigem Überlieferungskontext dienen. Schließlich werden die Forschungsergebnisse, insbesondere die Editionsprinzipien und das Metadatenschema, die Grundlage für weitere Fontane-Editionen bilden, etwa für die an der Fontane-Arbeitsstelle geplante erste digitale genetisch-kritische Edition des erzählerischen Werkes Theodor Fontanes.

Die fruchtbaren Erfahrungen im fortlaufenden Projekt zeigen schon jetzt, dass der Interaktion zwischen Editionsphilologie und Informationswissenschaft für das Gelingen des Projektziels eine zentrale Bedeutung zukommt. Die interdisziplinäre Zusammenarbeit führt vor allem dazu, dass beide Disziplinen wechselwirkend voneinander profitieren. So erlernen Editionswissenschaftler nicht nur den Umgang mit digitalen Arbeitsmethoden und -hilfsmitteln, sondern müssen beispielsweise erkennen, dass die Codierung handschriftlich überlieferter Aufzeichnungen eine noch genauere Transkription erfordert, damit die Befunde möglichst überlieferungsadäquat in ein präzises Mark-Up überführt werden können. Informationswissenschaftler hingegen gewinnen für weitere Arbeitszusammenhänge wertvolle Einblicke in die für

wissenschaftliche Text-Editionen notwendigen philologischen Methoden der Textkritik und Kommentierung von historischen handschriftlichen Aufzeichnungen und Drucken. Die Teamarbeit ermöglicht aber auch, dass, bedingt durch eine effektive Aufteilung, Fachwissenschaftler ihre professionellen Aufgaben nach wie vor wahrnehmen können. Die Virtuelle Forschungsumgebung ist schließlich auf vielfältige Projekte angewiesen, denn diese stellen nicht nur die Anforderungen, sondern sind vielmehr über Fehlermeldungen und Feedback am gesamten Entwicklungsprozess beteiligt. Genauso gilt dies für die Infrastrukturentwicklung im Allgemeinen, weshalb es notwendig ist, dass Forschungsprojekte auch direkt in diese Entwicklungen miteinbezogen werden. Dafür sind Kooperationen erforderlich, die gewährleisten, dass die damit verbundenen Synergie-Effekte effektiv genutzt werden können.

Im Unterschied zu traditionellen, in Buchpublikationen mündende Editionsverfahren ändert sich das *Procedere* bei der Realisierung digitaler Editionen insofern, als es mehr als zuvor das Material in den Mittel- und Ausgangspunkt der Arbeitsmethoden stellt: Vor der Festlegung eines Gesamtkonzepts und dem Beginn der Transkription steht eine systematische Autopsie des Materials, um die Beschreibung und Deutung aller Befunde exakt dokumentieren und in den Editionsprinzipien widerspiegeln zu können. Erst das Ergebnis dieser induktiven Vorgehensweise bildet die entscheidende Grundlage für das Metadatenschema und die Codierungsrichtlinien, die von einem Metadatenpezialisten in enger Zusammenarbeit mit den Philologen entwickelt werden. Die informationswissenschaftlich-technische Umsetzung geschieht abschließend wiederum in enger Kollaboration mit den jeweils anderen Disziplinen. Für die Konzipierung und Erarbeitung digitaler Editionen ist jedoch nicht nur die kooperative Zusammenarbeit einzelner Fachdisziplinen notwendig; vielmehr verschmelzen alle Wissenschaftsbereiche zu einer neuen angewandten und eigenständigen Wissenschaft, den Digital Humanities.

Literaturverzeichnis

Fontanes Notizbücher sind Eigentum der Staatsbibliothek zu Berlin, Preußischer Kulturbesitz, Handschriftenabteilung, Signatur: Nachl._Theodor_Fontane,Notizbücher. Für die Erlaubnis zur Abbildung danken wir herzlich.

Projektwebsite: <http://www.uni-goettingen.de/de/303691.html/>

- AFA – Gedichte (1989): *Theodor Fontane: Gedichte. Drei Bände*. Hrsg. von Joachim Krueger und Anita Golz. Berlin/Weimar: Aufbau.
- AFA – Romane und Erzählungen (²1973) [¹1969]: *Theodor Fontane: Romane und Erzählungen in acht Bänden*. Hrsg. von Peter Goldammer et al. Berlin/Weimar: Aufbau. *Band 3: Grete Minde, L'Adultera, Ellernklipp, Schach von Wuthenow*. Bearb. von Anita Golz und Gotthard Erler. *Band 5: Irrungen, Wirrungen, Stine, Quitt*. Bearb. von Jürgen Jahn und Peter Goldammer.
- AFA – Wanderungen (1976–1987): *Theodor Fontane: Wanderungen durch die Mark Brandenburg 1–5*. Hrsg. von Gotthard Erler und Rudolf Mingau. Berlin/Weimar: Aufbau.
- Berbig, Roland (2010): *Theodor Fontane Chronik*. Projektmitarbeit 1999–2004: Josefina Kitzbichler. 5 Bände. Berlin: de Gruyter.
- Busa, R. (1980): The Annals Of Humanities Computing: The Index Thomisticus. In: *Computers and the Humanities* 14 (1980), 83–90.
- Erler, Gotthard (2000): „Die Reisetagebücher“. In: Grawe, Christian; Nürnberger, Helmuth (Hrsg.): *Fontane-Handbuch*. Stuttgart: Kröner, S. 771 f.
- GBA (1994 ff.; noch nicht abgeschlossen): *Theodor Fontane: Große Brandenburger Ausgabe*. Begründet und hrsg. von Gotthard Erler. Fortgeführt von Gabriele Radecke und Heinrich Detering. Berlin: Aufbau. – *Das erzählerische Werk*. 21 Bände (1997 ff.). Hrsg. von Gotthard Erler, in Zusammenarbeit mit dem Theodor-Fontane-Archiv Potsdam. Editorische Betreuung: Christine Hehle. – *Gedichte*. 3 Bände (²1995). Hrsg. von Joachim Krueger und Anita Golz. – *Tage- und Reisetagebücher*. 3 Bände (1994–2012). Hrsg. von Charlotte Jolles, Gotthard Erler und Christine Hehle unter Mitarbeit von Therese Erler und Rudolf Muhs. – *Wanderungen durch die Mark Brandenburg*. 8 Bände (1997). Hrsg. von Gotthard Erler und Rudolf Mingau. Unter Mitarbeit von Therese Erler, Alfred Frank und Rita Reuter.
- HFA (1962–1997): *Theodor Fontane: Werke, Schriften und Briefe*. Hrsg. von Walter Keitel und Helmuth Nürnberger. München: Hanser. – *Abt. III, Band 3/II* (1997): *Erinnerungen, ausgewählte Schriften und Kritiken. Dritter Band: Reiseberichte und Tagebücher. Zweiter Teilband: Tagebücher*. Hrsg. von Helmuth Nürnberger et al.
- Hoffmann, Dierk O.; Zils, Harald (2005): „Hölderlin-Editionen“. In: Nutt-Kofoth, Rüdiger; Plachta, Bodo (Hrsg.): *Editionen zu deutschsprachigen Autoren als Spiegel der Editions-geschichte* (Bausteine zur Geschichte der Edition; Bd. 2). Tübingen: Niemeyer, S. 199–245.
- Jannidis, Fotis (2005): „Elektronische Edition“. In: Nutt-Kofoth, Rüdiger; Plachta, Bodo (Hrsg.): *Editionen zu deutschsprachigen Autoren als Spiegel der Editions-*

- geschichte* (Bausteine zur Geschichte der Edition; Bd. 2). Tübingen: Niemeyer, S. 457–470.
- Jannidis, Fotis (2008): Digitale Editionen. In: *Mitteilungen des Deutschen Germanistenverbandes* 55 (2008): 3, 317–332.
- Kamzelak, Roland S. (2012): Empfehlungen zum Umgang mit Editionen im digitalen Zeitalter. In: *editio* 26 (2012), 202–209.
- Kay, Michael (2007): XSL Transformations (XSLT) Version 2.0. <http://www.w3.org/TR/xslt20/>.
- Neuroth, Heike; Lohmeier, Felix; Smith, Kathleen Marie (2011): TextGrid – Virtual Research Environment for the Humanities. In: *The International Journal of Digital Curation* 6 (2011): 2, 222–231.
- NFA (1959–1975): *Theodor Fontane: Sämtliche Werke. Bd. 1–24 [und Supplemente]*. Hrsg. von Edgar Gross et al. München: Nymphenburger Verlagshandlung. – *Band XVIIIa* (1972): *Theodor Fontane: Unterwegs und wieder daheim*. Gesammelt von Kurt Schreinert, fortgeführt und hrsg. von Jutta Neuendorff-Fürstenau. – *Band XXIII/2* (1970): *Theodor Fontane: Aufsätze zur bildenden Kunst. Teil 2*. Gesammelt von Kurt Schreinert und Wilhelm Vogt, fortgeführt und hrsg. von Rainer Bachmann und Edgar Gross.
- Radecke, Gabriele (2002): *Vom Schreiben zum Erzählen. Eine textgenetische Studie zu Theodor Fontanes „L’Adultera“* (Epistemata; Bd. 358). Würzburg: Königshausen & Neumann.
- Radecke, Gabriele (2008): Theodor Fontanes Notizbücher. Überlegungen zu einer notwendigen Edition. In: Amrein, Ursula; Dieterle, Regina (Hrsg.): *Gottfried Keller und Theodor Fontane. Vom Realismus zur Moderne* (Schriften der Theodor Fontane Gesellschaft; Bd. 6). Berlin: de Gruyter, S. 211–233.
- Radecke, Gabriele (2010): Theodor Fontanes Notizbücher. Überlegungen zu einer überlieferungsadäquaten Edition. In: Schubert, Martin (Hrsg.): *Materialität in der Editionswissenschaft* (Beihefte zu *editio*; Bd. 21). Berlin: Niemeyer, S. 95–106.
- Radecke, Gabriele; de la Iglesia, Martin; Michaelis, Judith (2012): Genetisch-kritische und kommentierte Edition von Theodor Fontanes Notizbüchern. Vortrag, gehalten auf dem TextGrid Summit am 15. Mai 2012, Universität Würzburg. http://www.textgrid.de/fileadmin/TextGrid/konferenzen_vortraege/TextGrid_Summit_2012/TG-Summit_praesentation_fontane.pdf/.
- Saller, Harald (2003): HNML, HyperNietzsche Markup Language. In: *Jahrbuch für Computerphilologie* 5 (2003), 185–192. Online: <http://computerphilologie.digital-humanities.de/jg03/saller.html>.
- Schubarth-Engelschall, Angelika (1976): Die Notizbücher Theodor Fontanes. Eine Übersicht. In: *Fontane Blätter*, Sonderheft 4: Register für die Bände 1 (1965) bis 2 (1973) [Hefte 1–16] und die Sonderhefte 1–3 (1976), S. 64–66.

Die Edition „Johann Friedrich Blumenbach – online“ der Akademie der Wissenschaften zu Göttingen

von Martina Kerzel, Mike Reich und Heiko Weber¹

Das Projekt „Johann Friedrich Blumenbach – online“

Die Online-Edition der Schriften und naturhistorischen Sammlungen Johann Friedrich Blumenbachs (1752–1840; Abb. 1) verfolgt das Ziel einer „Gesamtschau“ und Analyse (d. h. Erschließung, Bearbeitung und Verfügbarmachung) von Blumenbachs Gesamtwerk. Dies soll durch Digitalisierung, Volltextfassung (der gedruckten Publikationen), Tiefenerschließung, und -indexierung sowie Kommentierung seiner Publikationen und der heute noch erhaltenen naturhistorischen Objekte aus seinen Sammlungen (und den Sammlungen des ehemaligen Kgl. Academischen Museums in Göttingen, die er betreute) erreicht werden. Das Ergebnis sind wissenschaftlich tiefenerschlossene Texte sowie Objekte, die für weiterführende Forschungsfragen, einzeln und verbindend, eine langzeitverfügbare und zitierfähige, digitale Grundlage für die Blumenbachforschung bilden. Das Besondere dieser Edition ist die Erstellung und Verknüpfung zwischen tiefenerschlossenen digitalisierten Texten und ebenso tiefenerschlossenen und digitalisierten Sammlungsobjekten.

Das bisher nur punktuell – und vorwiegend nach den Erkenntnisinteressen der Geisteswissenschaften, insbesondere der Wissenschaftshistoriker – ausgewertete Material kann dadurch nun im Gesamtkontext des 18. und frühen 19. Jahrhunderts – aus der Perspektive der Geistes- und Naturwissenschaften – erforscht werden. Zugleich entsteht mit dem dabei angestrebten Querschnitt durch alle Fachgebiete ein repräsentatives Abbild dieser Epoche des Übergangs von einer statisch verstandenen „Naturgeschichte“ zu einer tempo-

¹ Die hier vorliegenden Darstellungen der bisherigen Ergebnisse und Leistungen des Projekts „Johann Friedrich Blumenbach – online“ gründen sich auf die Arbeitsergebnisse der Mitarbeiter und Mitarbeiterinnen, der kooperativen Mitarbeiter und Mitarbeiterinnen und der Leitungskommission des Projekts. Nähere Informationen zur Zusammensetzung der Arbeitsstelle des Projekts unter <http://www.blumenbach-online.de/mitarbeiter/>.

realisierten „Geschichte der Natur“. Diese „verzeitlichte Sicht auf die Natur“ gilt es, mithilfe neuer, digitaler Verfahren in Beziehung zu den bisherigen Forschungsergebnissen und aktuellen Forschungsfragestellungen zu setzen.



Abb. 1 Johann Friedrich Blumenbach (1752–1840), zeitgenössische Radierung, angefertigt von Ludwig Emil Grimm (1790–1863) in Kassel (1823); Original im Geowissenschaftlichen Museum Göttingen

Die Virtuelle Forschungsumgebung (VRE)² TextGridLab und das Online-Portal „Johann Friedrich Blumenbach – online“ schaffen für das gesamte digitale Material des Projekts die Möglichkeit sowohl zu einer Volltextsuche über strukturelle und inhaltliche Aspekte als auch zu einer Suche auf Ebene der Metadaten (Texte und Objekte). Somit wird es den Wissenschaftlern und Wissenschaftlerinnen unterschiedlicher Forschungsbereiche erleichtert, Zugang zu den vielfältigen Quellen zu erlangen, Parallelen aufzudecken oder intertextuelle Bezüge offenzulegen. Darüber hinaus besteht die Möglichkeit, in dem geschützten Bereich der VRE (TextGrid) einen eigenen Forschungsbereich mit veröffentlichten Materialien der Blumenbachforschung anzulegen und eigenen Forschungsinteressen, sei es im Forschungsverbund oder einzeln, nachzugehen.

Das Editionsprojekt „Johann Friedrich Blumenbach – online“ der Akademie der Wissenschaften zu Göttingen, an dem die Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB) sowie das Geowissenschaftliche Zentrum der Universität Göttingen beteiligt sind, versteht sich hauptsächlich als ein interdisziplinäres Fach- und Themenprojekt für die geistes- und naturwissenschaftliche Forschung, an deren Schnittstelle die Informationswissenschaft steht.³ In diesem Zusammenhang ist nicht nur das Internet, sondern sind auch digitale Forschungsmethoden „aus dem Arbeitsalltag in den Geschichtswissenschaften nicht mehr wegzudenken“ (Burkhardt/Hohls/Prinz 2007: 11)⁴. Dabei hat vor allem auch die retrospektive Erschließung deutschsprachiger Textkorpora zurzeit Konjunktur, wenn nicht Hochkonjunktur.⁵ Neben den internationalen⁶ und den an deutschen Bibliotheken angesiedelten Digitalisierungsprojekten⁷ haben in den letzten Jahren auch Suchmaschinen-

2 Zur Definition einer Virtuellen Forschungsumgebung vgl. Lossau (2011). Einen generellen Überblick geben Carusi/Reimer (2010).

3 Zum Kontext der historischen Fach- und Themenportale vgl. u. a. Zündorf (2007), Schmale (2010), Gasteiner/Haber (2010) sowie Haber (2011).

4 Vgl. auch Pscheida (2010), Gleick (2011) und Brandstetter/Hübel/Tantner (2012).

5 Zu diesem Kontext vgl. u. a. Jacob (1996), Knoche/Tgahrt (1997), Koch (1998), Keller (2000), Horch/Schicketanz (2001), Dörr (2003), Keck (2004).

6 Bspw. die Bibliothèque nationale de France (s. <http://gallica.bnf.fr/>) oder das digitale Zeitschriftenarchiv informaworld (s. <http://www.informaworld.com/smpp/browse~db=all~thing=title~by=title>); eine Liste internationaler digitaler Zeitschriftenarchive ist verfügbar unter <http://onlinebooks.library.upenn.edu/serials.html>.

7 Vgl. <http://www.dfg.de/foerderung/>.

betreiber – wie bspw. Google Books⁸ – damit begonnen, zahlreiche Digitalisate von gemeinfreien historischen Monografien und Zeitschriften zu erstellen⁹ und über Suchmaschinen allgemein zugänglich zu machen (vgl. Mittler 1998; Thaller 2005).

Dabei sind Quellen aus dem 18. und frühen 19. Jahrhundert für die historische Forschung von besonderem Interesse. Denn in dieser Epoche geistlicher und weltlicher Neuorientierungen, in welcher Kommunikation und Austausch zunehmend komplexer und vor allem öffentlicher wurden (vgl. Frei 1999), war der Aufschwung der Publikationstätigkeiten eng mit gesellschaftlichen, politischen, sozialen und wissenschaftlichen Entwicklungen und Interaktionen verbunden. So entwickelten sich unterschiedliche Publikationsformen (vor allem die Zeitschriften und Journale) zum wichtigsten Transformationsmedium der alle Lebensbereiche umfassenden Debatten einer selbstreflexiv-aufgeklärten Gesellschaft (vgl. Greiling 2005).

Der Ausstrahlungs- und Rezeptionsraum der Handbücher, Zeitschriften und Journale erweiterte sich dabei schon in der Zeit vor 1800 (vgl. u. a. Böning 2005; Kopitzsch 2006), vor allem aber in der ersten Hälfte des 19. Jahrhunderts, über den Kreis der „Gelehrten“ hinaus auf nahezu alle sozialen Schichten (vgl. bspw. Pfeifer 1996; Kurzweg 1999).

Gerade die Vielzahl an Publikationen birgt somit ein enormes Potenzial im Hinblick auf die Erforschung der naturwissenschaftlichen, sozialen und politischen Strömungen dieser Zeit (vgl. Ries 2007): Sämtliche wichtigen Entwicklungen spiegeln sich hier wider und durch die Vielfalt an beteiligten Autoren (z. T. bekannte Schriftsteller, aber zunehmend auch engagierte Leser), Themen (z. B. ausländische Literatur, politische Nachrichten und philosophische oder naturwissenschaftliche Fragen) und Publikationsformen (wie bspw. wissenschaftliche Fachzeitschriften, moralische Wochenschriften oder auch Frauenblätter) ergibt sich ein repräsentatives Bild dessen, was die Gesellschaft des späten 18. und frühen 19. Jahrhunderts bewegte, wie sie ihr Weltbild formte und wie sich überkommene Vorstellungen im Zuge einer öffentlichen Debatte langsam ändern konnten. Jedoch bringt allein dieser exponentielle Zuwachs an Publikationen zu Beginn des 19. Jahrhunderts ein methodologisches Problem mit sich: Für den Geisteswissenschaftler wird es

8 Siehe <http://books.google.de/intl/de/googlebooks/about.html>.

9 Eine ausführliche Linkliste findet sich unter <http://www.kisc.meiji.ac.jp/~mmandel/recherche/digitalisierung.html>.

zunehmend schwieriger, angesichts des massiven Zuwachses an digital zugänglichen Quellenmaterial zu entscheiden, welche Quellen für sein spezifisch-methodologisches Erkenntnisinteresse relevant sind. Wissenschaftshistoriker sehen sich mit diesem Problem in geringerem Ausmaß konfrontiert, da zu Beginn des 19. Jahrhunderts eine Aufspaltung der Wissenschaften von einem universalen Ansatz hin zu einzelnen, sich immer stärker differenzierenden Fach- und Wissensdisziplinen einsetzte.¹⁰ Dies erlaubt eine Differenzierung und Strukturierung der Quellen, womit die gedruckten wissenschaftshistorischen Quellen durch ihre Spezialisierung selbst eine Eingrenzung für die wissenschaftshistorische Forschung geben.

Im Gegensatz zur Bedeutung dieses herausragenden, sich selbst eindeutig abgrenzenden Wissensfundus steht heute die Frage nach dem Grad der Zugänglichkeit und der Verfügbarkeit von gedruckten Quellen und Sammlungsobjekten, speziell für wissenschaftlich-komparatistische Fragestellungen. Bisher wurde bei den Erschließungsprojekten¹¹ der Bibliotheken vor allem darauf gesetzt, eine möglichst vollständige und umfangreiche Erschließung und Bereitstellung von gedruckten Schriften vorzunehmen. Digitalisierung und Erfassung von Sammlungsobjekten spielte in diesem Zusammenhang keine oder nur eine untergeordnete Rolle. Unter dem Gesichtspunkt der Kernaufgaben von Bibliotheken – Sammlung, Erschließung (formal, inhaltlich), Bewahrung und Bereitstellung von Informationsquellen (vgl. Hobohm

10 Ein Beispiel sind die *Annalen der Physik*, die seit über 200 Jahren zu den bedeutendsten Fachzeitschriften der wissenschaftlichen Literatur gehört. Sie entstand 1799 als Nachfolgerin einer Reihe von Zeitschriften, die durch den deutschen Chemiker Friedrich Albrecht Carl Gren herausgegeben worden waren. Dieser publizierte von 1790 bis 1794 zunächst das *Journal der Physik*, dem später – von 1795 bis 1797 – das *Neue Journal der Physik* folgte. Nach seinem Tod wurde diese Zeitschriftenreihe mit dem bis heute gebräuchlichen Titel *Annalen der Physik* von Ludwig Wilhelm Gilbert (Professor an der Universität Halle) fortgesetzt. Die Zeitschrift besitzt eine besondere wissenschaftshistorische Bedeutung, weil in ihr eine Vielzahl von bahnbrechenden wissenschaftlichen Erkenntnissen publiziert wurde. Vgl. Schimank (1963).

11 Vgl. das Projekt „Index deutschsprachiger Zeitschriften 1750–1815“ (abgeschlossen) bzw. „Gelehrte Journale und Zeitungen als Netzwerke des Wissens im Zeitalter der Aufklärung“ (laufend) der Akademie der Wissenschaften zu Göttingen sowie das von der Universitätsbibliothek Bielefeld in Zusammenarbeit mit dem Georg Olms Verlag aufgelegte Projekt „Retrospektive Digitalisierung wissenschaftlicher Rezensionenorgane und Literaturzeitschriften des 18. und 19. Jahrhunderts aus dem deutschen Sprachraum“.

2012) sowie deren aktive Vermittlung (vgl. Plassmann et al. 2006: 8f.) – findet dieses Verfahren durchaus seine Rechtfertigung. Vor allem in Hinsicht auf wissenschaftlich-komparatistische Fragestellungen der historisch arbeitenden Wissenschaften stößt diese Erschließung der gedruckten Quellen – wie das Beispiel Google Books augenfällig zeigt – aber an ihre Grenzen. Von wenigen Ansätzen abgesehen fehlt bisher ein interdisziplinäres Arbeitsinstrument für die wissenschaftliche Forschung, welches über eine Gesamtschau der gedruckten Quellen eine kontextbezogene Einordnung der thematisierten Fragestellungen verschiedener Disziplinen ermöglicht.

Vorteile der Nutzung von Digitalisaten und Volltexten im Internet

Die nachweisbaren Exemplare der Publikationen und die Sammlungsobjekte J. F. Blumenbachs sind auf mehrere (z. T. auch ausländische) Bibliotheken und museale Sammlungen verteilt; einige seltene Schriften können nur vor Ort in den entsprechenden Bibliotheken eingesehen werden. Museale Objekte sind per se nur in den Sammlungen, in denen sie verwahrt werden, verfügbar. Die Online-Edition stellt diese Materialien (Schriften und vor allem auch die naturhistorischen Sammlungsobjekte) – soweit überliefert und noch vorhanden – vollständig digital zur Verfügung und gewährleistet der weltweiten Forschung einen orts- und zeitunabhängigen Zugriff auf die Quellen. Ebenso bietet die Online-Edition die Möglichkeit zur fortlaufenden Aktualisierung des editierten Werkskorpus (Texte und Objekte).

Durch den freien Zugriff auf Digitalisate und Volltexte via Internet beseitigt die Online-Edition zudem sozio-ökonomische Barrieren (wie bspw. Reisekosten, Kosten für die Anfertigung von Digitalisaten etc.) und schont die originalen Schriften und Objekte vor unnötiger Abnutzung. Des Weiteren erleichtert der Zugriff auf die Texte und Sammlungsobjekte von Blumenbach via digitalen Informationstechnologien den Zugang zu entsprechenden Informationen auch für Menschen mit körperlichen Einschränkungen (wie bspw. sehschwache und blinde Personen).

E-Research in „Blumenbach – online“: Bearbeitung bestehender Forschungsfragen mit weiteren Mitteln

Die Online-Edition ermöglicht die variable Präsentation der Werkgenese von Blumenbachs Schriften und Sammlungsobjekten. Der Forscher oder die For-

scherin hat damit die Freiheit, die für seine Forschung relevanten und notwendigen Textfassungen gezielt auszuwählen und selbst zu entscheiden, ob er einen dia- oder synchronen Vergleich der Texte vornehmen möchte. Durch die Verknüpfung von Texten und Objekten, Beschreibungen (Metadaten) und Visualisierungen wird eine Analyse von Blumenbachs Gesamtwerk z. B. mit quantitativen Methoden ermöglicht. Zudem bleiben sowohl Analysegrundlage wie auch -methode und -ergebnisse zitierbar, verfügbar und validierbar. Im Gegensatz zu einer gedruckten Edition (vgl. Plachta 2006) kann der Forscher persönlich bestimmen, wie er die Verknüpfung von Texten, Beschreibungen, Abbildungen und Visualisierungen (3D-Animationen, Geo-Browsing etc.) vornehmen möchte. Somit besteht die Möglichkeit einer variablen, nutzerspezifisierten Präsentation der Inhalte der Online-Edition und der auf ihr basierenden Forschungsergebnisse (z. B. in Hinsicht des Zusammenhangs zwischen neuen empirischen Erkenntnissen und Definitionen bei Blumenbach).

Komfortable, personalisierbare Funktionen (wie Browsen, Suchen, Filtern, schneller Zugriff auf Datenbestände und auf die in einem gewählten Kontext stehenden Informationen) ermöglichen nicht nur das serielle Blättern und Erschließen in der Blumenbach-Edition (wie in einer gedruckten Ausgabe), sondern erlauben auch die freie Wahl der Art des Zugriffes. Der Forscher trifft selbst die Entscheidung darüber, ob er die Originalquellen (= *Quellenansicht*) und/oder die Edition der Herausgeber (= *Editionsansicht*) bemüht, ob er seriell ganze Texte lesen möchte oder sich via Filter und Suchfunktionen vergleichbare Textabschnitte zugänglich machen möchte (z. B. zum Vergleich von Versionen inhaltlich und sprachlich verwandter Texte mithilfe des Synopsentools). Im Gegensatz zu einer Druckversion, die nur eine einzige und statische – durch die Herausgeber festgelegte – Präsentationsform der Inhalte erlaubt, stellt die Online-Edition einen dynamischen Prozess der Generierung und Präsentation der Quellen- und Editions-korpora dar. Visualisierungen von Text- und Objektdaten – bzw. von Beziehungen zwischen Daten – können deshalb nach spezifischen Nutzer- bzw. Forschungsinteressen generiert werden (vgl. Sahle 2013).

Das In-Bezug-Setzen (Verlinken, Annotieren und Kommentieren) von Texten und Objekten in der Online-Edition ermöglicht es, Bedeutungsinhalte von Texten und Objekten aufzuweisen. So wird z. T. sowohl der Sinn als auch die Bedeutung (im Sinne von Frege [1892]) eines in einem Text benannten Objektes erst durch den Rückgriff auf das dem Text zugrunde liegende Sammlungsobjekt erkennbar. Umgekehrt erhält der Text in der Zuord-

nung der Objekte einen höheren Grad an „Sinnlichkeit“. Texte können damit nicht nur klarer, sondern vor allem auch deutlicher werden (im Sinne von *clare et distincte*)¹².

Darüber hinaus erlaubt die Online-Edition einen vereinfachten Zugriff auf die Datenbestände zur Erhebung von biometrischen (bspw. anthropologischen), statistischen, prosopografischen etc. Daten sowie für Forschungen innerhalb verschiedener Disziplinen. Vor allem das „Rollen“-Argument der TEI-XML-Auszeichnung der Volltexte erlaubt eine prosopografische Netzwerkanalyse in Hinsicht auf die Genese und Etablierung von Blumenbachs Werken und Wirken.¹³

Der Zugang via TextGridLab ermöglicht den Aufbau von und die Teilhabe an einem internationalen Wissensnetzwerk der fachübergreifenden Blumenbachforschung. Die Auszeichnung der Volltexte (vgl. Espen 2012; Schmidt 2012) mit TEI-XML, P5, Best Practice (BP) Level 5¹⁴ ermöglicht die Erweiterung der Auszeichnung der Blumenbachtexte um weitere fachspezifische Kontexte (z.B. linguistische und philologische Forschungen).¹⁵ Bei TEI handelt es sich um einen auf XML basierenden Auszeichnungsstandard der Text Encoding Initiative¹⁶, der in den Geisteswissenschaften, insbesondere in den Editionswissenschaften und der Linguistik, verbreitet ist. Die Entscheidung für diesen (und weitere) Standards und die Nutzung desselben (und weiterer) Standards konnte und kann mit Unterstützung der Fachkolle-

12 Zum Begriff von „*clare et distincte*“, der auf die Bestimmungen von René Descartes in seinen *Meditationes de prima philosophia* beruht, vgl. u. a. *Meditationes de prima philosophia* (lateinischer Originaltext von 1641, englische Übersetzung von 1901, französische Übersetzung von 1647) (s. <http://www.wright.edu/cola/cartes/>) sowie den Eintrag „Klarheit“ in Rudolf Eislers *Wörterbuch der philosophischen Begriffe* (s. <http://www.textlog.de/3798.html>).

13 Zum Begriff der Netzwerke vgl. Fangerau/Halling (2009).

14 „Best Practices for TEI in Libraries“, erstellt durch die SIG TEI in Libraries (s. <http://www.tei-c.org/SIG/Libraries/teiinlibraries/>)

15 Zur Umsetzung dieser Ziele wurde durch die Akademie der Wissenschaften zu Göttingen (AdWG) und der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW) ein Kooperationsvertrag abgeschlossen, der zum Ziel hat, alle deutschsprachigen Monografien von Johann Friedrich Blumenbach im Deutschen Textarchiv (DTA) der BBAW als TEI-XML, P 5, BP Level 3, bis Ende 2013 zur Verfügung zu stellen. Details zum DTA: <http://www.deutschestextarchiv.de>.

16 Vgl. <http://www.tei-c.org/index.xml>.

ginnen und -kollegen aus dem Bereich der Informationswissenschaft (insbesondere der SUB Göttingen) erfolgen.

Des Weiteren gewährleistet das Trierer Wörterbuchnetzwerk (bisher 24 Wörterbücher¹⁷) im TextGridLab den Zugriff des Forschers oder der Forscherin auf weitergehende Informationen im Kontext des Blumenbachschen Werks (Erläuterungen, Erklärungen, Übersetzungen etc.). Von besonderem Interesse sind hier das Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm (DWB); das Goethe-Wörterbuch (GWB) sowie die *Oekonomi-sche Encyklopaedie* von Johann Georg Krünitz (Kruenitz). Die Einbindung des für das Blumenbachprojekt einschlägigen, da zeitgenössischen Wörterbuchs (Ersch-Gruber) „Allgemeine Encyclopädie der Wissenschaften und Künste in alphabetischer Folge von genannten Schriftstellern bearbeitet und herausgegeben von J[ohann] S[amuel] Ersch und J[ohann] G[ottfried] Gruber (Professoren zu Halle). Mit Kupfern und Charten. Leipzig, im Verlage von Johann Friedrich Gleditsch zwischen 1818 und 1886, in 3 Sektionen (A–G, H–Ligatur, O–Phyxios) in insgesamt 167 Bänden mit ca. 70.000 Seiten“ in das Trierer Wörterbuchnetz wurde durch das Blumenbachprojekt vorgeschlagen und ist zeitnah vorgesehen.

E-Research in „Blumenbach – online“: Generierung neuer Forschungsfragen

Vor allem die Nutzung des Synopsentools via TextGridLab und TextGridRep wird die Generierung von neuen Forschungsfragen erlauben. Durch die elektronisch unterstützte Ermittlung von Gemeinsamkeiten und Unterschieden von Textvarianten (unterschiedliche Ausgaben und Auflagen eines Werkes, dessen Übersetzung in unterschiedliche Sprachen, unterschiedliche Gattungen von Texten zu einem Thema etc.) können für den Forscher Sachverhalte sichtbar gemacht werden, die mit großer Wahrscheinlichkeit zu neuen Forschungsfragen führen. Vor allem die Differenzialanalyse der von Blumenbach innerhalb der Genese der Schriften verwendeten Fachbegriffe, Nomenklaturen und Systematisierungen ist hier von großem Interesse. Durch die Auszeichnung der Fachbegriffe durch TEI-XML in den Volltexten stehen diese für statistisch-vergleichende, digitale Analyseverfahren wie z.B. Text-

17 Siehe <http://woerterbuchnetz.de>.

mining¹⁸ zur Verfügung und erlauben es, über Such- und Filterfunktionen die im Fokus der Forschung stehenden Begriffe – nebst Kontext in variablem Ausmaß – zu lokalisieren, zu exportieren und diese mittels weiterer Werkzeuge anderer Anwendungssoftware einer statistischen Analyse zu unterziehen. Von Interesse sind hier z. B. besonders die Variantenerschließung und der Wandel von lateinischen zu deutschen Fachbegriffen. So wäre beispielsweise über die Volltexte einfacher zu untersuchen, wie Blumenbach die Objektklasse der „Naturalia“ vielfach auf einen gemeinsamen und eindeutigen Begriff bringt.

Die große Herausforderung besteht darin, zum einen Forschungsfragen zu formulieren, welche die wissenschaftshistorischen und naturwissenschaftlichen Fachgrenzen überwinden können, sowie gleichzeitig die Daten so aufzubereiten, dass diese Fragen beantwortet und die Daten nachhaltig interpretierbar präsentiert werden können.

Quellen und angewandte Methoden

Gegenstand (Quellen) des Projekts sind einerseits Blumenbachs eigene Publikationen und direkt aus diesen abgeleitete Veröffentlichungen (zumeist Übersetzungen) sowie andererseits die noch vorhandenen Objekte aus Blumenbachs naturhistorischen Sammlungen. Auch veröffentlichte und unveröffentlichte Briefe (sowie dazugehörige Briefregesten)¹⁹ stellen wichtige Bindeglieder zwischen Publikationen und Sammlungsobjekten dar, wobei diese bei letzteren oftmals eine genaue Provenienzanalyse ermöglichen, insbesondere wenn begleitende Originaletiketten bei Objekten fehlen oder die dortigen Angaben unvollständig sind.

Blumenbach selbst publizierte auf Latein und Deutsch; übersetzt wurden seine Werke ins Englische, Französische, Italienische, Spanische, Russische, Dänische und Niederländische. Das Projekt wird zukünftig unterstützend handschriftliches Material heranziehen, z. B. Blumenbachs annotierte Handexemplare, sein handschriftliches Verzeichnis zur Reiseliteratur oder seine entsprechenden Sammlungskataloge. Die Blumenbachschen Sammlungsob-

18 ausführliche Erläuterungen zum Verfahren des Text Mining u. a. in Hausser (2000), Mehler/Lobin (2004); Heyer/Quasthoff/Wirrig (2008), Carstensen (2010), Lobin (2010)

19 Siehe http://www.blumenbach-online.de/fileadmin/wikiuser/Daten_Digitalisierung/Briefregesten/Blumenbach_Briefregesten.html.

jekte umfassen das gesamte Spektrum der Naturgeschichte seiner Zeit und sind heute Gegenstand so unterschiedlicher Fachdisziplinen wie Anatomie, Anthropologie, Medizin, Archäologie, Ethnografie, Botanik, Pharmakognosie, Zoologie, Paläontologie, Mineralogie und Geologie (vgl. Reich et al. 2012: 162 f.). Objekte und Texte bilden hierbei eine Einheit: einerseits in Form des Gegenstandsbereichs – Texte, als deskriptive Ebene der Empirie, und Sammlungsgegenstände als eine aufgrund von Befundanalysen getroffene und validierte Auswahl der in Göttingen (und auswärtig) erhalten gebliebenen Sammlungsobjekten aus dem ehemaligen Göttinger Kgl. Academischen Museum und Blumenbachs Privatsammlungen (vgl. Reich/Gehler 2012: 170 ff.) – sowie andererseits deren Kontexte, also Methoden und Darstellungsformen des Beschreibens, Erläuterns, Erklärens und Verstehens bei Blumenbach.

Zentral ist hier die Frage, wie das Verhältnis von Idee und Theorie zu den empirischen Befunden (in seinen Kontinuitäten, Diskontinuitäten und Interdependenzen) in einer „Einheit in der Vielheit“, also in den in der Online-Edition publizierten und tiefererschlossenen Texten (Deskription) in Rückgriff auf die identifizierten und tiefererschlossenen Sammlungsobjekte (Empirie) im Gesamtwerk von Blumenbach steht. Wesentliches Ziel des Projekts ist deshalb auch, die Blumenbachschen Sammlungsobjekte in einem virtuellen Online-Museum wieder zusammenzuführen, mit dessen Hilfe es dann möglich ist, aus den Objekten explizite Verweise auf die Blumenbachtexte (und umgekehrt) für die Forschung präsentieren und zugänglich machen zu können.

Texterfassung, Digitalisierung und Tiefererschließung

Bis Ende 2012 wurde die Digitalisierung fast aller (99%) in der Bibliografie zu Johann Friedrich Blumenbach verzeichneten gedruckten Schriften abgeschlossen (vgl. Kroke 2010). Aus den Digitalisaten (TIFF-Scans) wurden PDF-Dateien generiert, welche der öffentlichen Nutzung zur Verfügung gestellt wurden.²⁰

²⁰ Diese Daten sind online (s. http://www.blumenbach-online.de/fileadmin/wikiuser/Daten_Digitalisierung/Bibliographie/Bibliographie.html) verfügbar und werden 2013 in das Online-Portal (Forschungsportal) „Johann Friedrich Blumenbach – online“ integriert.



<pre> 172 <pb n="[3]" face="images/00000003" xml:lang="de" xml:id="pb003_0001"/> 173 <figure xml:lang="de" xml:id="003a"> 174 <graphic url="images/00000707_003a"/> 175 <head rendition="center">Bildschöner Schedel einer Georgianerin.</ head> 176 <figDesc>Textabbildung.</figDesc></figure> 177 <pb n="interleaf" face="images/00000004" xml:lang="de" xml:id="pbinterleaf_0006"/> 178 <pb n="[5]" face="images/00000005" xml:lang="de" xml:id="pb005_0001"/> 179 <head rendition="center">51.-</div>Bildschöner Schedel einer Georgianerin.</head> 180 <p>Ein Menschenschedel - sey's auch von welchem Volke unter der Sonne es wolle - verhalten mit dem vom aller-menschenähnlichsten Affen, ist allein schon hinreichend, den körperlichen Character der Humanität zu bestimmen, und die Elufst zu ermessen, welche die Natur zwischen menschlicher und thierischer Gestaltung befestigt hat. Und so finden zwey zusammenfassende Abbildungen dieser beiderley sich so sehr von einander ausreichenden Schädelformen auch in dieser Sammlung eine passende Stelle. Der von einem Neger würde zwar außerdem auch zugleich gelehrt haben, mit einem Blicke zu zeigen, wie selbsterdings bloße relative die Züge sind, wodurch er sich von den Köpfen anderer Menschensaffen unterscheidet; und wie gewöhnlich er längere durchsicht in denen was nur am Schedel den constanten abschätzen Character der Humanität ausmacht, mit allen übrigen Völkern abzuzeichnen, und so get wie irgend eines derselben, von der übrigen thierischen Schöpfung absetzt. - Ich habe aber doch lieber hier diesen gewählt, weil er wohl allgemeineres Interesse, wegen der bewundernswürdigen Schönheit seiner Bildung haben wird, wodurch er sich so auffallend von mehr als hundert andern Schädeln fremder Völkerschleichen auszeichnet, wozu nun meine Sammlung angesehen ist. Er ist eine der vielen und großen Geschenke, womit der Herr Baron von Aetz diese Sammlung bereichert hat, und dient zur Beglaubigung dessen, was außer so vielen andern Zeugen von der Schönheit der Georgianer, namentlich der philosophische Reisende, Cassini, von den bezaubernden Reizen ihrer Weiber sagt *). * Voyage en Perse etc. T. I. pag. 171. der Amsterd. Aug. 1735. 4. 87/16138. </pp></note></p></div> <div data-bbox=</pre>
--

Verknüpfung von Text-Digitalisat, zugehörigem Volltext; ggf. Link zu Sammlungsobjekt

Kennzeichnung einer Überschrift

originale Worttrennung (im elektron. Volltext ggf. ausblendbar)

Personennamen und Verknüpfung mit einer Personen-Datenbank

Information über die Verwendung von Kapitälchen im Originaltext

Original-Fußnote und Verweis auf eine Literaturdatenbank; ggf. Hyperlink zu externer online-Bibliothek

Abb. 2 Aufbereitung von Digitalisaten im Projekt „Johann Friedrich Blumenbach – online“. Oben: Historische Bild und Textseiten. Unten: digitaler Volltext der Publikationen mit TEI-XML-kodierten Zusatzinformationen z. B. zu Textgestalt und -struktur, Textverständnis und mit Hyperlinks zu Zusatzangeboten. (Abb. Wolfgang Böker: Projekt „Johann Friedrich Blumenbach – online“)

Die Aufbereitung der Blumenbach-Texte folgt den Ebenen (1) Inhaltserschließung, (2) Korpus-immanente Analyse und (3) Korpus-externe Kontextualisierung. Mit der Inhaltserschließung wird das Wissen, welches Blumen-

bach bei seinen Zeit- und Fachgenossen implizit voraussetzen konnte, rekonstruiert und in die heutige Begrifflichkeit übersetzt. Erreicht wird dies in erster Linie durch die Identifikation von Personen, Orten, Ereignissen und Objekten.²¹

Historische oder ungebräuchlich gewordene Bezeichnungen werden durch aktuelle ergänzt. Die Korpus-immanente Analyse sucht nach textgenetischen Zusammenhängen innerhalb des Blumenbachschen Werks, während die Korpus-externe Kontextualisierung die Beziehungen der Blumenbach-Texte zu Texten anderer Autoren dokumentiert. Ausgangspunkt sind dabei Blumenbachs eigene Verweise auf die von ihm verwendete Literatur. Ausgezeichnet wird vor allem in Hinsicht auf Personennamen, Körperschaftsnamen, Ortsnamen, Literaturangaben, Zitate, Fachbegriffe und Erläuterungen, Sammlungsobjekte, Abbildungsbeschreibungen, redaktionelle Anmerkungen, Datumsangaben, typografischen Buchschmuck, Sprachwechsel, Auflösungen von Abkürzungen, unsichere Lesungen und nicht auflösbare Unklarheiten, Streichungen, Hinzufügungen und Ergänzungen des Autors (hier vor allem bei den Schriften mit handschriftlichen Anmerkungen von Blumenbach), Druckfehler, Normalisierungen und fehlende Texte.

Zur Disambiguierung des personen- und ortsbezogenen Datenmaterials wird kontrolliertes Vokabular verwendet: Für die dafür notwendige Einbindung persistenter Identifier (PID) wird in Hinsicht auf Personen- und Körperschaftsnamen der CERL-Thesaurus²² genutzt. CERL steht für „Consortium of European Research Libraries“. Es handelt sich um einen Zusammenschluss europäischer wissenschaftlicher Bibliotheken, die u. a. den CERL-Thesaurus pflegen, der auch Personennamen verzeichnet. Der Vorteil der CERL-Nutzung liegt darin, dass anders als bei der GND²³ (die im Übrigen mit dem CERL-Thesaurus verknüpft ist) auch Nicht-Autoren nachgewiesen werden können. Projekte wie „Blumenbach – online“, die mit CERL arbeiten, bekommen ein eigenes Kürzel zugewiesen, das jeden durch das Projekt angelegten Datensatz zu einer Person als projektmotiviert identifiziert. Darüber hinaus enthält jeder Datensatz einen CERL-Thesaurus-eigenen Persistent Identifier, der die Daten nachhaltig referenzierbar macht. Für geografische Ortsbezeichnungen wird der über TextGrid eingebundene „Getty-Thesaurus

21 Das Blumenbach-Projekt folgt somit dem im DDB-Kontext erarbeiteten Who-what-when-where-Modell.

22 Siehe <http://thesaurus.cerl.org/cgi-bin/search.pl?start=true>.

23 Siehe http://www.dnb.de/DE/Standardisierung/Normdaten/GND/gnd_node.html.

of Geographic Names“²⁴ benutzt. Diese Datenbank systematisch strukturierter Ortsbezeichnungen enthält neben historischen Bezeichnungen und Ortstypen auch die geografischen Koordinaten der Orte (Angaben zu geografischen Längen- und Breitengraden) und macht sie auf diese Weise eindeutig identifizierbar – solange man sie im vorliegenden auszuzeichnenden Text eindeutig identifizieren kann.

Der Nutzen von Standardvokabular bzw. Thesauri liegt in dem nachhaltig disambiguierten Datenmaterial, das wissenschaftlich tiefenerschlossen und durch die Verknüpfung mit anderen Daten einen Mehrwert generiert. Dieses Datenmaterial kann zukünftig mithilfe von beispielsweise Visualisierungstools Fragestellungen beantworten, u. a. mithilfe des Geobrowsers²⁵ die Fragen, in welchem Zeitraum sich bestimmte Ideen Blumenbachs über welchen geografischen Raum ausgebreitet haben oder welche Einzelstationen Sammlungsobjekte in welchem Zeitraum vom Fundort bis zum Untersuchungs- oder Ausstellungsort zurückgelegt haben.

Sammlungserfassung, Digitalisierung und Tiefenerschließung

Nach Lokalisierung, Verzeichnung und Identifikation von Blumenbachschen Sammlungsobjekten (s. Abb. 3) widmet sich der Arbeitsbereich Sammlungserfassung, Digitalisierung und Tiefenerschließung vor allem der Erschließung und Digitalisierung naturhistorischer Objekte. Bei ausgewählten Objekten kommt neben der fotografischen 2D-Dokumentation das halbautomatisierte 3D-Digitalisierungsgerät PackshotCreator 3D (inkl. des Zusatzmoduls PackshotViewer) der Fa. Sysnext (Paris) zur Anwendung (Abb. 3).

Dabei werden von geeigneten Objekten bis zu einer maximalen Größe von $40 \times 40 \times 40$ cm bis zu 240 Einzelaufnahmen für eine Halbkugel-Animationsdarstellung erzeugt; bei besonderen bzw. geeigneten Objekten können auch Vollkugel-Animationen erzeugt werden (vgl. Reich et al. 2012: 163 f.). Zur nachhaltigen Langzeitarchivierung dieser Einzeldateien werden sie zusätzlich zum JPG-Format im hochauflösenden, farbgetreuen, nicht komprimierten RAW-Format gespeichert. Um den 3D-Effekt im Browser anzeigen zu können, werden aus diesen Einzeldateien Animationen im Flash-, GIF- oder HTML5-Format (bevorzugt letzteres) erzeugt. Zusätzlich dazu besteht

24 Siehe <http://www.getty.edu/research/tools/vocabularies/tgn/index.html> (02.04.2013).

25 Vgl. <http://dev2.dariah.eu/e4d/> (02.04.2013).

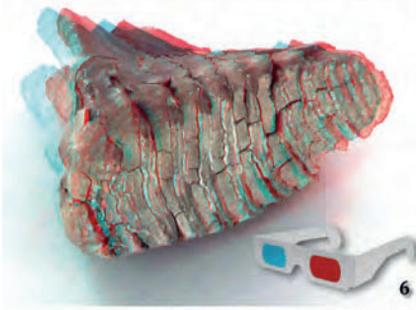
(bei Flash oder HTML5) die Möglichkeit, in diese 3D-Animationen hinein-zuzoomen. In Abhängigkeit von der Auflösung (normalerweise 300 dpi) zzgl. der Erstellung der Animation (s.o.) muss pro Objekt (3D-Digitalisierung) von einer Gesamt-Datenmenge von 150 MB bis zu 3 GB ausgegangen werden. Gemäß der Empfehlungen der DFG²⁶ werden aus den RAW-Dateien TIFF-Dateien generiert, um die Langzeitverfügbarkeit der Bilddaten bestmöglich zu gewährleisten.

Größere Objekte (> 40 cm Durchmesser) werden bei Eignung zusätzlich zur herkömmlichen 2D-Digitalisierung (vgl. ebd.) teilweise mit einer Stereokamera digitalisiert (FinePix Real 3D W1 von Fuji), die Anaglyphenbilder mit stereoskopischem Effekt erzeugt. Bei der 2D-Digitalisierung muss pro Objekt mit einer Datenmenge von bis zu 50 MB gerechnet werden, wobei jedoch im Internetportal nur kleinere, heruntergerechnete Dateien zu sehen sein werden (die „Masterfiles“ werden im RAW-Format und als TIFF nachhaltig gesichert), die bei Bedarf im Hintergrund geladen werden und für eine Detailansicht zur Verfügung stehen.

In vorangegangenen Arbeitsschritten wurden umfangreiche fachspezifische (Anatomie, Anthropologie, Medizin, Archäologie, Botanik, Pharmakognosie, Zoologie, Paläontologie, Mineralogie, Geologie etc.) Metadaten-schemata für die Sammlungsobjekte entwickelt. Sie bilden jeweils die Grundlage für die Weiterentwicklung hinsichtlich: (1) Texte (Zusammenarbeit mit dem Arbeitsbereich Texterfassung, Digitalisierung und Tiefenerschließung) und (2) Datenbank (Zusammenarbeit mit dem Arbeitsbereich VRE und Online-Portal). Archivarbeiten im Hinblick auf die Sammlungsobjekte wurden bisher in den einzelnen Göttinger Sammlungsinstitutionen wie auch im Universitätsarchiv und in der SUB Göttingen durchgeführt. Diese werden im Zuge der Arbeiten noch auf andere Archive (z.B. Hauptstaatsarchiv Hannover) ausgedehnt.

Blumenbachs multidisziplinäres Schaffen spiegelt sich in der Vielfalt der naturhistorischen (und teilweise fächerübergreifenden) Objekte in seinen Sammlungen wider. Bisher konnten 15 naturwissenschaftliche Fachrichtungen (nach heutiger Fachsystematik) determiniert werden, in welche die Sammlungsobjekte einzuordnen sind. Damit umfasst diese Zuordnung (historisch und nach modernen naturwissenschaftlichen Ordnungssystematiken und Klassifikationen) fast alle modernen naturwissenschaftlichen Fächer.

26 standardisiert lt. Vorgaben bzw. Empfehlungen der DFG-Praxisregeln „Digitalisierung“ 2013: http://www.dfg.de/download/pdf/foerderung/programme/lis/praxisregeln_digitalisierung_2013.pdf



Benagte Knochen
aus der Hya(e)meahöhle
bey Kirkdale in Yorkshire
v. H. Prof. Buckland in Oxford



Benagte Knochen
aus der Hya(e)meahöhle
bey Kirkdale in Yorkshire
v. H. Prof. Buckland in Oxford

8

9

Abb. 3 (1–4) Lokalisierung, Verzeichnung und Identifikation von naturhistorischen Objekten aus den ehemaligen Privatsammlungen J. F. Blumenbachs bzw. aus dem von ihm mehr als 60 Jahre lang betreuten *Kgl. Akademischen Museum (Blumenbachiana)*. (1) Blick in die paläontologische Hauptsammlung des Geowissenschaftlichen Zentrums der Universität Göttingen (GZG), Kompaktusanlage mit mehr als 27.000 Schubladen; (2) Blick in einen Regalbereich der Wirbeltierpaläontologischen Sammlung im GZG; (3) Blick in die Teilsammlung „Historische Geowissenschaften“ mit tlw. ausgezogener Schublade, heute einen Großteil der noch vorhandenen *Blumenbachiana* beherbergend; (4) Beispiele für verschiedene *Blumenbachiana* (aus den Bereichen Botanik, Geologie, Meteoritenkunde und Ethnografie) mit assoziierten Primär-, Sekundär- oder Tertiäretiketten. (5–7) Digitalisierung von *Blumenbachiana*. (5) Herkömmliche 2D-Digitalisierung unter Zuhilfenahme einer digitalen Spiegelreflexkamera (nebst Farbskala); (6) ‚Einfache‘ 3D-Digitalisierung unter Zuhilfenahme einer digitalen Stereokamera und einer aus den Aufnahmen erzeugten Anaglyphenaufnahme (nebst zur Betrachtung notwendigen Rot/Cyan-Brille aus Leichtkarton); (7) Beispiele aus einer Serie von bis zu 240 Einzelaufnahmen, die mit dem halbautomatisierten 3D-Digitalisierungsgerät PackshotCreator 3D (Fa. Sysnext, Paris; unten rechts im Bild) angefertigt wurden. Aus diesen Einzelaufnahmen können nachfolgend am Computer Halbkugel- bzw. Vollkugel-Animationsdarstellungen erzeugt werden. (8–9) Tiefenerschließung von *Blumenbachiana*. (8) Transkriptionen, Transliterationen (tlw. auch Übersetzungen) von den Objekten beiliegenden (oder aufgeklebten) Primär-, Sekundär-, Tertiär- oder Quartäretiketten; oft ist neben dem Etikett Blumenbachs auch das Originaletikett des Sammlers / Donators / Verkäufers erhalten; (9) Überprüfung, inwieweit die untersuchten Objekte oder Serien Belegstücke oder Abbildungsoriginale zu einer seiner zahlreichen Veröffentlichungen darstellen, oder in seiner Korrespondenz Erwähnung finden.

(alle Fotos und Objekte: Geowissenschaftliches Museum Göttingen bzw. Projekt „Johann Friedrich Blumenbach – online“ an der Göttinger Akademie der Wissenschaften)

Zur Aufnahme der Metadaten wurde von den Projektmitarbeitern des Arbeitsbereichs „Objektdigitalisierung“ ein Metadatenkonzept erstellt. Anhand dieses Schemas wurde ein Datenbankmodell entwickelt und eine webbasierte grafische Nutzeroberfläche (GUI) programmiert, die online ein globales und paralleles Arbeiten ermöglicht. Pro Objektsuite ist es möglich, über 120 unterschiedliche Metadaten zu erfassen. Eingebettet in den internen Website-Bereich des Projekts (<http://www.blumenbach-online.de>) ist das GUI. Die Authentifizierung des Zugriffs auf das Web-GUI zur Eingabe der Daten erfolgt mittels des Typo3-CMS, welches auch für die Pflege der Webseiten verwendet wird. Zur Datenspeicherung wird eine MySQL-Datenbank verwendet.

Die Metadaten werden, soweit möglich, standardisiert eingegeben. Zum Beispiel ist in das Erfassungs-GUI der Objekt-Datenbank ein Zugang zum CERL-Thesaurus für Personen und zum Getty-Thesaurus für die eindeutige Identifizierung von Ortsbezügen der Objekte und Serien integriert, welcher vom eingebenden wissenschaftlichen Mitarbeiter komfortabel über ein Auswahlmenü bedient werden kann und neben der entsprechenden Personen-normdaten-ID (CERL-ID) bzw. der Ortsnamen-ID (Getty-ID) auch automatisch die Geokoordinaten bzw. die standardisierten Orts- und Personennamen ermittelt und erfasst.

Die Integration der mittels der Erfassungsmaske gesammelten MySQL-Daten in die Virtuelle Forschungsumgebung (VRE) TextGridLab wurde ebenfalls bereits programmiert. Weiterhin wurde ein weitgehend automatisiert stattfindender Workflow erarbeitet, um die in der Sammlungsmaske erfassten Daten aus der MySQL-Datenbank in ein XML-Format zu konvertieren, sodass die Daten in die eXist-Datenbank des TextGridRep „ingestiert“ (übertragen, integriert, importiert) werden können.

Auch eine webbasierte Präsentation der Sammlungsobjekt-Daten aus TextGridLab zurück in die Webseiten des Blumenbachprojekts ist in Arbeit. Hierfür wird das an der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW) entwickelte SADE-Framework nachgenutzt.²⁷ SADE arbeitet im Kern mit einer XML-Datenbank (eXist) und ermöglicht, Transformationsszenarien (XSLT, XQuery, XSL-FO) auszuführen und das Ergebnis über eine Webschnittstelle abzufragen. Mittels dieser Schnittstelle ist es möglich, die in der VRE TextGridLab vorhandenen und wissenschaftlich aufbereiteten Sammlungsobjekte und Schriften Blumenbachs nach beliebigen Fragestellungen aufbereitet auszugeben (inklusive aller Objektverknüpfungen untereinander), im Internet durchsuchbar zu machen und zur Verfügung zu stellen (z. B. innerhalb des Blumenbach-Portals). Ein Demonstrator ist bereits implementiert und wird kontinuierlich weiterentwickelt.

Virtuelle Forschungsumgebung und Online-Portal

TextGridLab steht seit dem 07.05.2012 in der Version 2.0 zur Verfügung. Seitdem wurden weitere, für das Projekt „Blumenbach – online“ spezifische Anpassungen der Virtuellen Forschungsumgebung selbst vorgenommen bzw.

²⁷ Vgl. <http://www.bbaw.de/telota/projekte/digitale-editionen/sade/sade-1> (02.04.2013).

werden derzeit realisiert, von denen die beiden wesentlichen Aspekte kurz dargestellt werden: Anpassungen des Ingest-Workflows sowie Anpassungen der Metadaten schemata.

Der Ingest-Workflow für Objektdaten und zugehörige Bilddaten inkl. Versionierung (vgl. Abb. 4) wurde angepasst, um einen Abgleich der erfassten Daten zu den Sammlungsobjekten mit wachsender Tiefenerschließung zu gewährleisten. Hierdurch wird sichergestellt, dass die erfassten Daten nach einer Einspeisung in das TextGrid-Repository aktualisiert werden können, ohne dass sich die persistente Identifizierung ändert. Auf diese Weise bleiben die Bezüge zwischen Objekt-Digitalisaten, Text-Digitalisaten, Daten und Texten ebenfalls persistent. Hier übernimmt das Projekt „Blumenbach – online“ eine Vorreiterrolle, denn bisher nutzen vornehmlich ausschließlich textbasierte Editionsprojekte die Arbeitsumgebung von TextGrid. Bei diesen stellt sich das beschriebene Problem nicht, da für sie vor allem eine persistente Verbindung zwischen Digitalisat und Volltext (aber nicht zu Objekten) erforderlich ist.

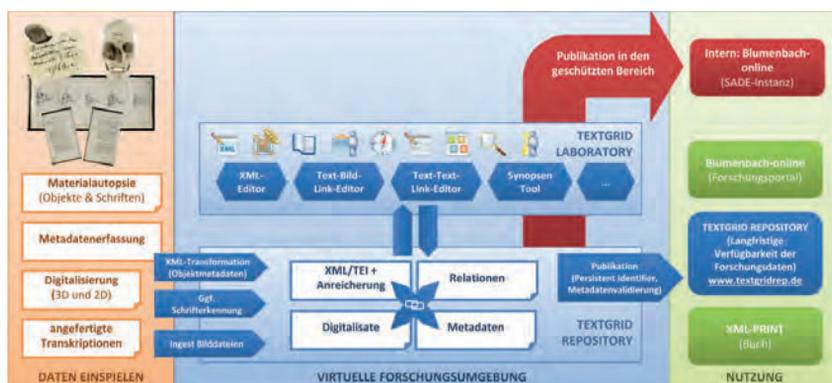


Abb. 4 Ingest-Workflow der Text- und Sammlungsdaten des Projekts „Johann Friedrich Blumenbach – online“ in TextGridLab und TextGridRep und deren möglicher Bearbeitung mit TextGridTools in TextGridLab (Abb. TextGrid/Martina Kerzel: Abteilung Forschung & Entwicklung, SUB Göttingen)

Das in TextGrid verwendete und für eine übergreifende Suche notwendige Metadaten-Kernset wurde deshalb um objektspezifische Angabemöglichkeiten erweitert, sodass die Suche und Findbarkeit von Sammlungsobjekten in der TextGrid-Umgebung optimal sichergestellt ist. Bisher bezogen sich die Metadatenelemente des Kernmetadaten schemas in TextGrid auf textba-

sierte Elemente. Dementsprechend wurden beim Import vor der Version 2.0 von TextGrid nur Daten zu Autor und Titel angefragt. Für die Behandlung von Sammlungsobjekten wurde mit der Version 2.0 ein weiteres, standardisiertes Datenschema erarbeitet, welches Angaben zu Sammlungsobjekten – bspw. Angaben zum Donator – weiterverarbeiten kann.

Weitere Spezifikationen betreffen zwei Werkzeuge, die prädestiniert waren, auf besondere Anforderungen des Projekts „Blumenbach – online“ angepasst zu werden. Es handelt sich hierbei um den Text-Text-Link-Editor sowie um das Synopsentool bzw. einen auf die spezifischen Anforderungen des Projekts zugeschnittenen Kollationierer (weiterhin – wie bisher – als Synopsentool bezeichnet).

Die Programmierung des Text-Text-Link-Editors²⁸ wurde durch das TextGrid-Team der FH Worms realisiert. Dazu hat das Blumenbachprojekt dem TextGrid-Konsortium seine spezifischen Anforderungen mitgeteilt. Diese werden mit den Anforderungen aus weiteren Fachbereichen zusammengeführt, damit auf diese Weise ein möglichst generisch einsetzbares Werkzeug entsteht, das eine 1:n-Beziehung zwischen unterschiedlichen Volltexten bzw. Textstellen unterstützt. Inwieweit die spezifischen Ausprägungen, die seitens des Projekts „Blumenbach – online“ erforderlich sind, von der generischen Umsetzung des Tools abweichen und ob sich eventuell weiterer Anpassungsbedarf für die VRE „Blumenbach – online“ ergibt, wird sich im konkreten Arbeitsprozess innerhalb der VRE ergeben.

Ein ebenfalls Blumenbach-spezifisches Werkzeug innerhalb der VRE stellt das bereits erwähnte Synopsentool dar. Dieses Werkzeug ermöglicht den Vergleich mehrerer Textpassagen und die automatische Auswertung; wahlweise werden Übereinstimmungen oder Abweichungen herausgestellt. Das Synopsentool wird in seiner speziellen Ausprägung für die Blumenbach-spezifischen Anwendungen auf das Kollationierungs-Tool in TextGridLab Version 2.0 und das CollateX-Tool²⁹ aufbauen. Insbesondere wissenschaftshistorische Erkenntnisse, die z. B. auf dem Vergleich mehrerer aufeinanderfolgender Ausgaben eines Blumenbach-Werkes beruhen, lassen sich auf diese Weise von Nutzern der digitalen Edition innerhalb und außerhalb der VRE nachvollziehen.

28 Siehe <https://dev2.dariah.eu/wiki/display/TextGrid/Text+Text+Link+Editor>.

29 Siehe <https://dev2.dariah.eu/wiki/display/TextGrid/CollateX>.

Zurzeit werden umfangreiche Tests innerhalb der VRE mit dem nun zur Verfügung stehenden Material durchgeführt. Die Daten haben zwar ihr komplexes Endstadium bei weitem noch nicht erreicht, dennoch sind die derzeitigen Testläufe und Weiterentwicklungen des Workflows wichtig, um etwaige Schwierigkeiten bzw. Lücken im Workflow bereits frühzeitig zu erkennen und Maßnahmen zu Datenharmonisierung und Ingest in die VRE erfassen, analysieren und an die blumenbachspezifischen Erfordernisse anpassen zu können. Die Testläufe umfassen sowohl gedruckte Werke (Digitalisate, Volltexte, Metadaten) als auch Erfassung, Aufbereitung und Ausgabe der Daten zu den Sammlungsobjekten (Digitalisate, Metadaten, Tiefenerschließung).

Für diejenigen Werke, die aus dem Bestand der SUB Göttingen stammen, existiert bereits ein Workflow³⁰, der die automatische Weiterverarbeitung aller technischen und strukturellen Daten (METS/MODS)³¹ gewährleistet, was u. a. für die Langzeitarchivierung wichtig ist. Beim Import in das TextGridRep werden diese Daten unverändert übernommen und weitere Daten, je nach gewählten Konfigurationseinstellungen, hinzugefügt, z. B.:

- TextGrid URI (Uniform Resource Identifier) zur dauerhaften und eindeutigen Identifizierbarkeit von Objekten innerhalb des TextGridLab (verbunden mit einem eindeutigen PID z. B. für die Langzeitarchivierung bei der GWDG);
- automatische Aktualisierung von bestehenden Verbindungen zwischen Objekten unter Berücksichtigung der beim Import erstellten URIs. Objekte können mithilfe der in TextGrid zur Verfügung stehenden Versionierung nach einem Import- und Exportverfahren erneut importiert werden, ohne dass ein neuer URI vergeben wird, d. h. alle erteilten Verlinkungen bleiben erhalten.

Unberücksichtigt blieb bisher der Ingest von nicht an der SUB Göttingen erstellten Digitalisaten. Damit diese in den automatisierten Workflow integriert und ebenfalls mit allen z. B. für die Langzeitarchivierung notwendigen technischen Daten ausgestattet werden können, muss der am Göttinger Digitalisierungszentrum (GDZ) übliche Workflow – namentlich die Ausstattung der Digitalisate mit METS/MODS-Daten – integriert werden. Hierbei kann

30 Göttinger Digitalisierungszentrum: Workflow im GDZ; online: <http://gdz.sub.uni-goettingen.de/gdz/workflow-im-gdz>

31 Siehe <http://gdz.sub.uni-goettingen.de/entwicklung/standardisierung>.

das Projekt auf die Expertise der Abteilung Metadaten und Datenkonversion³² zurückgreifen, die diesen Prozess fachgerecht durchführt.

Die nächsten wichtigen Schritte sind der Aufbau der Blumenbach-Edition innerhalb der VRE, um mit der Verknüpfungs- und Analysearbeit zwischen Volltexten, Textdigitalisaten und Sammlungsobjekten (Digitalisate und Metadaten) beginnen zu können. Vorbereitende Maßnahmen in diesem Zusammenhang waren bspw. für den Bereich der Sammlungsobjekt-Datenbank die Bereitstellung der Exportschnittstelle mit einer Funktionalität, welche die exportierten Daten mit einer für den Ingest in das TextGridLab notwendigen Konfigurationsdatei anreichert.

Ein erster Prototyp des Online-Portals „Johann Friedrich-Blumenbach – online“ (in Form eines „Demonstrators“) wurde bereits im November 2011 realisiert. Für das Portal werden derzeit ein Konzept sowie verschiedene, granulare Nutzungs- und Anwendungsszenarien erarbeitet und erprobt. Zu berücksichtigen ist hierbei die anspruchsvolle Heterogenität des Datenmaterials, das nach der erfolgten Verknüpfung innerhalb der Präsentationsschicht der VRE sowie im Online-Portal den Wissenschaftlern zur Verfügung gestellt werden soll. Der seit November 2011 betriebene Demonstrator basiert auf dem an der BBAW entwickelten SADE-Framework. Der Export der Daten aus dem TextGridLab in das Online-Portal erforderte einige projektspezifische Anpassungen:

- Implementierung von jQuery-Abfragen über die REST-Schnittstelle (Representational State Transfer), welches das vorhandene Datenmaterial so aufbereitet, dass die gewünschte Repräsentationsform im Demonstrator erreicht wird.
- Implementierung einer Löschfunktion für Aggregationen innerhalb der SADE-Instanz. Hierzu wurde ein Xquery-Skript erstellt, welches mittels SPARQL-Abfrage den TG-CRUD-Service abfragt und die zu löschenden Dateien ermittelt und löscht.
- Implementierung einer speziellen Datenaufbereitung für die Präsentationsschicht. Dies verlagert die Rechenlast vom Nutzer-Rechner auf den SADE-Server. Nutzerseitig sind nur noch wenig rechenintensive Renderingprozesse notwendig. Die Darstellung erfolgt dabei weitgehend mit jQuery und HTML.

32 Siehe <http://www.sub.uni-goettingen.de/kontakt/abteilungen-a-z/abteilungs-und-gruppendetails/abteilunggruppe/metadaten-und-datenkonversion/>.

- Implementierung des DigiLib-Tools³³. DigiLib wird in der neuen Version des Demonstrators für die Präsentation und die automatische Skalierung aller Bilder genutzt. Mithilfe dieses Tools können beispielsweise Bildausschnitte markiert und als URL versendet werden.

Derzeit sind auf dem StorNext-System des Rechenzentrums der Universität (GWDG, Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen) 15 Terabyte zur Datensicherung des Projekts verfügbar, die per inkrementellem Backup auf den Bandroboter des Rechenzentrums gesichert werden. Sobald alle Daten in das TextGridRep überführt worden sind, sollen diese dort ebenfalls langzeitarchiviert werden.³⁴

Zusammenfassung und Ausblick

Eine Online-Edition wie „Johann Friedrich Blumenbach – online“ führt über die bisher etablierten Arbeitsgebiete im Bereich Online-Editionen durch die spezifischen Zielsetzungen des Projektes hinaus. Die Verknüpfung von tiefererschlossenem Textmaterial mit tiefererschlossenem Sammlungsmaterial (mit zugehörigen Metadaten, Normierungen, Normalisierungen und Standardisierungen) bietet Herausforderungen, an denen das Projekt – vor allem in der inter- und transdisziplinären Zusammenarbeit zwischen verschiedenen Fachwissenschaften wie Wissenschaftsgeschichte, Lebenswissenschaften, Geowissenschaften und Informationswissenschaft – innovative Grundlagenarbeit und Expertise (vor allem in Hinsicht auf die praktische Anwendung von Regelwerken) leistet.

Von besonderem Interesse in der derzeitigen Projektarbeit sind die bisher nur eingeschränkt verfügbaren Normierungen und Standardisierungen für die Repräsentation digital verfügbarer naturwissenschaftlicher Sammlungen und Objektdigitalisierungen. Hier musste das Projekt bisher eigene Wege gehen und wird in der weiteren Arbeit die verschiedenen Schemata der Textobjekte und der Sammlungsobjekte aufeinander mappen. Jedoch existieren in der Erschließung und Digitalisierung von Objektgattungen in wissenschaftlichen Sammlungen bereits verschiedene vielversprechende Ansätze (z. B. LIDO³⁵

33 Vgl. <http://www.mpg.de/410783/forschungsSchwerpunkt1>.

34 Vgl. TextGrid <http://www.textgrid.de/ueber-textgrid.html>.

35 Siehe <http://network.icom.museum/cidoc/working-groups/data-harvesting-and-interchange/lido-technical/specification/>.

oder aber der Metadata-Crosswalk aus dem Infrastrukturprojekt DARIAH-DE³⁶), die in Hinsicht auf konkrete Anwendungsprofile (application profiles) in der konkreten Sammlungserfassung validiert werden müssen. Deshalb prüft das Projekt in Zusammenarbeit mit Fachwissenschaftlern (u. a. von der SUB Göttingen und den Universitätsbibliotheken der Universität Göttingen, der Koordinierungsstelle für wissenschaftliche Universitätsbibliotheken Berlin sowie dem Deutsches Museum München), welche der derzeit verfügbaren Lösungen den größten Mehrwert und Anwendungsmöglichkeiten für das Projekt und die wissenschaftshistorische Forschung birgt.

Ein weiterer wichtiger Aspekt des Projekts ist die Nachhaltigkeit und somit dauerhafte Nutzung der Online-Edition „Johann Friedrich Blumenbach – online“. Einerseits ist das Projekt selbst auf Nachhaltigkeit angewiesen: Beispielhaft genannt werden hier die frei zur Verfügung stehende Software TextGridLab und der kostenfreie Storage Service mit Langzeitarchivierungsdiensten³⁷ (BitStream-Technology via Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen – GWDG) des TextGridRep. Es bestehen seitens der Förderer von TextGrid die Forderung und die Bemühungen, die Nachhaltigkeit dieser im Projektstatus befindlichen Angebote langfristig zu gewährleisten.

Das Projekt „Johann Friedrich Blumenbach – online“ der Akademie der Wissenschaften zu Göttingen selbst hat das Ziel, die Edition „Blumenbach – online“ nachhaltig und dauerhaft zur Verfügung zu stellen. Das Projekt ist somit bis zu einem gewissen Grad auf die technische und organisatorische Nachhaltigkeit der genutzten Komponenten, wie z. B. das TextGridRep, angewiesen. Hier hat die SUB als verlässliche Infrastruktureinrichtung ihre Rolle zur dauerhaften Bereitstellung von digitalen Daten bereits übernommen und wird die digitale Edition „Blumenbach – online“ dauerhaft nutzbar zur Verfügung stellen. Mit der Gründung des TextGrid-Vereins – dessen Aufgabe es ist, gemeinsam mit Förderinstitutionen und politischen Vertretern für eine dauerhafte Organisationsstruktur der TextGrid-Komponenten zu sorgen – ist ein wichtiger Meilenstein auf dem Weg in die Nachhaltigkeit gesetzt.

36 Vgl. Homepage von DARIAH-DE: <http://www.de.dariah.eu/>.

37 Zur konzeptuellen und technischen Dimension digitaler Bestandspflege vgl. Tibbo (2012). Zur Rolle der Langzeitarchivierung in den digitalen Geisteswissenschaften vgl. Gladney (2012).

Durch die Nutzung von Standards und die Auszeichnung der Quellen in XML ist bereits ein wichtiger Baustein in Richtung technischer Nachhaltigkeit gelegt. Mit diesem universalen und erweiterbaren Austauschformat bleiben die Quellen unabhängig von proprietär genutzter Software.

Eine Herausforderung, vor der viele Projekte stehen, wenn sie Quellen aus verschiedenen Einrichtungen (auch international) virtuell zusammenführen, sind urheberrechtliche Aspekte. Nutzungsrechte an Digitalisaten werden im Regelfall zu bestimmten Bedingungen großzügig (ohne Einschränkungen) für die wissenschaftliche Nutzung im Projektkontext erteilt. Einschränkende Bedingungen und Nutzungsrechte der liefernden Einrichtungen sind jedoch selbstverständlich vollumfänglich umzusetzen, auch wenn dies bedeutet, dass bei der Nutzung eines Objektes innerhalb des „Blumenbach – online“-Portals der Nutzer unter Umständen von der Portalseite weg- und auf die Seite der anbietenden Einrichtung hingeführt wird. „Blumenbach – online“ selbst stellt seine Forschungsergebnisse unter Open-Access-Bedingungen zur Verfügung, eingeschränkt in denjenigen Fällen, in denen das Projekt an bestimmte Nutzungsrechte digitaler Objekte gebunden ist. Dies für den Nutzer transparent im Portal umzusetzen, ist eine der bevorstehenden Aufgaben, die alle Fachwissenschaftler gemeinsam mit Unterstützung der in Urheberrechtsfragen erfahrenen Juristen (bspw. innerhalb der AG Elektronisches Publizieren der Akademie der Wissenschaften zu Göttingen) und der SUB Göttingen lösen werden.

Die aktive (ausschließliche) Nutzung des TextGridLab zur Erstellung der Edition „Johann Friedrich Blumenbach – online“ ist für die unmittelbare Zukunft geplant. Mittlerweile sind die Digitalisate, Volltexte und ein erheblicher Teil der Sammlungsobjekte aus Blumenbachs Dissertation *De generis humani varietate nativa* (vgl. Kroke 2010)³⁸ (deutschsprachige Ausgabe: *Über die natürlichen Verschiedenheiten im Menschengeschlechte*) (vgl. ebd.)³⁹ sowie aus seinem *Handbuch der Naturgeschichte* (vgl. ebd.)⁴⁰ in das TextGridLab bzw. TextGridRep überführt. Die Verknüpfungsarbeit zwischen Digitalisaten, Volltexten, Abbildungen, digitalisierten Sammlungsobjekten

38 online verfügbar unter: http://www.blumenbach-online.de/fileadmin/wikiuser/Daten_Digitalisierung/ID%2000001/00001.pdf

39 online verfügbar unter: http://www.blumenbach-online.de/fileadmin/wikiuser/Daten_Digitalisierung/Bibliographie/Bibliographie.html#00007

40 online verfügbar unter: http://www.blumenbach-online.de/fileadmin/wikiuser/Daten_Digitalisierung/Bibliographie/Bibliographie.html#00020

und Sammlungsobjekt-Metadaten kann somit beginnen. Für diesen Arbeitsprozess wird zurzeit ein (teil-) automatisierter Workflow entwickelt.

Ein nächster Schritt in dem nun anstehenden digitalen Forschungsprozess ist die Identifikation von beschriebenen Objekten in Blumenbach-Texten und die eindeutige Zuordnung des digitalisierten und tiefererschlossenen Sammlungsobjekts, auf deren Grundlage eine Verlinkung erfolgen kann. Hierzu ist durch das Projekt „Johann Friedrich Blumenbach – online“ die Nutzung und die Erweiterung des TextGridLab um projektspezifische Werkzeuge oder auch die Erweiterung des Publikationsframeworks SADE vorgesehen. Dies wird ermöglicht, weil TextGrid einerseits durch den Open-Source-Charakter der eigenen Entwicklungen und andererseits durch die modulare Architektur und die verwendeten Standards einen Anknüpfungspunkt für projektbezogene Entwicklungen beinhaltet.

Das Portalkonzept und die Repräsentation der Daten werden mit den Anforderungen der Nutzer und den sich entwickelnden technologischen Möglichkeiten innerhalb der nächsten zwölf Jahre (bis zum Projektende 2024) wachsen. Die Kooperation zwischen Fachwissenschaften und Bibliothek (insbesondere mit der Abteilung Forschung & Entwicklung der SUB Göttingen) hat sich bereits in den letzten drei Jahren als sehr fruchtbar und gewinnbringend erwiesen und das Projekt „Johann Friedrich Blumenbach – online“ der Akademie der Wissenschaften zu Göttingen wird auch in den kommenden Jahren in Fragen der VRE und des Online-Portals (Nachhaltigkeit, Organisation und Technologie) und darüber hinaus von dieser intensiven Zusammenarbeit profitieren.

Literaturverzeichnis

- Böning, Holger (2005): Bremer Zeitungen im 18. und frühen 19. Jahrhundert. In: Blome, Astrid; Böning, Holger (Hrsg.): *Täglich neu! 400 Jahre Zeitungen in Bremen und Nordwestdeutschland*. Bremen: Schünemann, S. 40–47.
- Brandstetter, Thomas; Hübel, Thomas; Tantner, Anton (Hrsg.) (2012): *Vor Google. Eine Mediengeschichte der Suchmaschine im analogen Zeitalter*. Bielefeld: transcript Verlag.
- Burkhardt, Daniel; Hohls, Rüdiger; Prinz, Claudia (Hrsg.) (2007): *Geschichte im Netz: Praxis, Chancen, Visionen. Beiträge der Tagung – hist2006*. Teilband I. (*Historisches Forum* 10, I, 2007). [http://edoc.hu-berlin.de/histfor/10_I/PDF/ Editorial_2007-10-I.pdf](http://edoc.hu-berlin.de/histfor/10_I/PDF/Editorial_2007-10-I.pdf).

- Carstensen, Kai-Uwe; Ebert, Christian; Ebert, Cornelia; Jekat, Susanne; Klabunde, Ralf; Langer, Haegen (Hrsg.) (2010): *Computerlinguistik und Sprachtechnologie. Eine Einführung*. 3. überarb. u. erw. Auflage. Heidelberg: Spektrum Akademischer Verlag.
- Carusi, Annamaria; Reimer, Torsten: Virtual Research Environment – Collaborative Landscape Study. A JISC funded project (January 2010). www.jisc.ac.uk/media/documents/publications/vrelandscape-report.pdf.
- Dörr, Marianne (2003): Digitale Bibliothek – Retrospektive Digitalisierung in Bibliotheken – vom Projekt zum Programm. In: *Zeitschrift für Bibliothekswesen und Bibliographie*, 84, 309–318.
- Fangerau, Heiner; Halling, Thorsten (2009): *Netzwerke: Allgemeine Theorie oder Universalmetapher in den Wissenschaften? Ein transdisziplinärer Überblick*. Bielefeld: transcript Verlag.
- Frege, Gottlob (1892): Über Sinn und Bedeutung. In: *Zeitschrift für Philosophie und philosophische Kritik, N.F.*, 100, 25–50.
- Frei, Andreas (1999): Die publizistische Bühne der Aufklärung: Zeitschriften als Schlüssel zur Öffentlichkeit. In: Stammen, Theo (Hrsg.): *Kant als politischer Schriftsteller*. Würzburg: Ergon, S. 29–42.
- Gasteiner, Martin; Haber, Peter (2010): *Digitale Arbeitstechniken für Geistes- und Sozialwissenschaften*. Wien/Köln/Weimar: Böhlau.
- Gladney, Henry M. (2012): Long-Term Digital Preservation: A Digital Humanities Topic? In: Thaller, Manfred (Hrsg.): *Controversies around the Digital Humanities (Historical Social Research – Historische Sozialforschung, HSR; 37/3)*, S. 201–217.
- Gleick, James (2011): *Die Information. Geschichte, Theorie, Flut*. Aus dem amerikanischen Englisch von Almuth Braun. München: Redline.
- Greiling, Werner (2005): Zeitschriften und Verlage bei der Vermittlung bürgerlicher Werte. In: Hahn, Hans Werner (Hrsg.): *Bürgerliche Werte um 1800: Entwurf – Vermittlung – Rezeption*. Köln: Böhlau, S. 211–224.
- Haber, Peter (2011): *Digital Past. Geschichtswissenschaften im digitalen Zeitalter*. München: Oldenbourg.
- Hausser, Roland (2000): *Grundlagen der Computerlinguistik. Mensch-Maschine-Kommunikation in natürlicher Sprache*. Berlin et al.: Springer.
- Heyer, Gerhard; Quasthoff, Uwe; Wirrig, Thomas (2008): *Text Mining: Wissensrohstoff Text. Konzepte, Algorithmen, Ergebnisse*. 1. korr. Nachdruck. Herdecke/Bochum: W3L.

- Hobohm, Hans-Christoph (2012): Can Digital Libraries Generate Knowledge? In: Thaller, Manfred (Hrsg.): *Controversies around the Digital Humanities (Historical Social Research – Historische Sozialforschung, HSR; 37/3)*, S. 218–229.
- Horch, Hans Otto; Schicketanz, Till (2001): „Ein getreues Abbild des jüdischen Lebens“. Compact Memory – ein DFG-Projekt zur retrospektiven Digitalisierung jüdischer Periodika im deutschsprachigen Raum. In: *Menora. Jahrbuch für deutsch-jüdische Geschichte*, 12, 387–409.
- Jacob, Herbert (1996): Retrospektive Zeitschriftenererschließung. In: *Informationsmittel für Bibliotheken*, 4, 610–621.
- Keck, Rudolf W. (2004): *Bildung im Bild: Bilderwelten als Quellen zur Kultur- und Bildungsgeschichte; ein Beitrag im Zusammenhang mit dem DFG-Projekt Retrospektive Digitalisierung ausgewählter Bibliotheksbestände*. Bad Heilbrunn/Obb.: Klinkhardt.
- Keller, Alice (2000): Berichte – Retrospektive Digitalisierung von Zeitschriften. In: *Nachrichten für Dokumentation*, 51 (2000): 4, 243–245.
- Knoche, Michael; Tghart, Reinhard (Hrsg.) (1997): *Retrospektive Erschließung von Zeitschriften und Zeitungen: Beiträge des Weimarer Kolloquiums, Herzogin-Anna-Amalia-Bibliothek, 25. bis 27. September 1996*. Berlin: Deutsches Bibliotheksinstitut (*Informationsmittel für Bibliotheken, IFB; Beiheft 4*).
- Koch, Michael (1998): Retrospektive Digitalisierung von Bibliotheksbeständen. In: *AKMB-News*, 4 (1998): 2–3, 24–25.
- Kopitzsch, Franklin (2006): Zeitungen – „Das Idol von Hamburg“: Zeugnisse zur Zeitungslektüre der Hamburger im 18. Jahrhundert. In: *Jahrbuch für Regionalgeschichte*, 24, S. 101–114.
- Kroke, Claudia (unter Mitarbeit von Reimer Eck und Wolfgang Böker) (2010): *Johann Friedrich Blumenbach – Bibliographie seiner Schriften*. Göttingen: Universitätsverlag (Schriften zur Göttinger Universitätsgeschichte; 2). Online: http://www.blumenbach-online.de/fileadmin/wikiuser/Daten_Digitalisierung/Bibliographie/Bibliographie.html.
- Kurzweg, Martina (1999): *Presse zwischen Staat und Gesellschaft. Die Zeitungslandschaft in Rheinland-Westfalen (1770–1819)*. Paderborn: Schöningh.
- Lobin, Henning (2010): *Computerlinguistik und Texttechnologie*. Paderborn: Wilhelm Fink.
- Lossau, Norbert (2011): Virtuelle Forschungsumgebungen und die Rolle von Bibliotheken. In: *Zeitschrift für Bibliothekswesen und Bibliographie*, 58, S. 156–165. Online: http://zs.thulb.uni-jena.de/servlets/MCRFileNodeServlet/jportal_derivate_00210988/j11-h3-4-auf-10.pdf.

- Mehler, Alexander; Lobin, Henning (Hrsg.) (2004): *Automatische Textanalyse. Systeme und Methoden zu Annotation und Analyse natürlichsprachlicher Texte*. Wiesbaden: VS Verlag für Sozialwissenschaften/GWV Fachverlage GmbH.
- Mittler, Elmar (Hrsg.) (1998): *Retrospektive Digitalisierung von Bibliotheksbeständen: Berichte der von der Deutschen Forschungsgemeinschaft einberufenen Facharbeitsgruppen „Inhalt“ und „Technik“*. Berlin: Dt. Bibliotheksinstitut (dbi-materialien; 166). Online: http://webdoc.sub.gwdg.de/pub/sub/2007/mittretr_PPN516494147.pdf.
- Ore, Espen S. (2012): Document Markup – Why? How? In: Thaller, Manfred (Hrsg.): *Controversies around the Digital Humanities (Historical Social Research – Historische Sozialforschung, HSR; 37/3)*, S. 106–124.
- Pfeifer, Hans-Wolfgang (1996): *Zeitungslandschaft: Prozesse, Profile, Perspektiven*. München: Koehler & Amelang.
- Plachta, Bodo (2006): *Editionswissenschaft. Eine Einführung in Methode und Praxis der Edition neuerer Text* (Reclams Universal-Bibliothek; Nr. 17603). 2., ergänzte u. aktualisierte Auflage. Stuttgart: Philipp Reclam jun.
- Plassmann, Engelbert; Rösch, Hermann; Seefeldt, Jürgen; Umlauf, Konrad (2006): *Bibliotheken und Informationsgesellschaft in Deutschland. Eine Einführung*. Wiesbaden: Harassowitz.
- Pscheida, Daniela (2010): *Das Wikipedia-Universum. Wie das Internet unsere Wissenskultur verändert*. Bielefeld: transcript.
- Reich, Mike; Gehler, Alexander (2012): Der Ankauf der Privatsammlung von J. F. Blumenbach (1752–1840) durch die Universität Göttingen. In: *Philippia* 15 (2012): 3, 169–187.
- Reich, Mike; Böhme, Annina; Gehler, Alexander; Numberger-Thuy, Lea D. (2012): „Preziosen jeglicher Couleur“ – Objektdigitalisierung der naturhistorischen Sammlungen von Johann Friedrich Blumenbach (1752–1840). In: *Philippia* 15 (2012): 2, 155–168.
- Ries, Klaus (2007): *Wort und Tat: das politische Professorentum an der Universität Jena im frühen 19. Jahrhundert* (Pallas Athene; 20). Stuttgart: Steiner.
- Sahle, Patrick (2013): *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Teil 1: Das typografische Erbe; Teil 2: Befunde, Theorie und Methodik; Teil 3: Textbegriffe und Recodierung* (Schriften des Instituts für Dokumentologie und Editorik; Bde. 7–9). Norderstedt: Books on Demand
- Schimank, Hans (1963): Ludwig Wilhelm Gilbert und die Anfänge der „Annalen der Physik“. In: *Sudhoffs Archiv für Geschichte der Medizin und Naturwissenschaften*, 47, S. 360–372.

- Schmale, Wolfgang (2010): *Digitale Geschichtswissenschaft*. Wien/Köln/Weimar: Böhlau.
- Schmidt, Desmond (2012): The Role of Markup in the Digital Humanities. In: Thaller, Manfred (Hrsg.): *Controversies around the Digital Humanities (Historical Social Research – Historische Sozialforschung, HSR; 37/3)*, S. 125–146.
- Thaller, Manfred (Red.) (2005): Retrospektive Digitalisierung von Bibliotheksbeständen: Evaluierungsbericht über einen Förderschwerpunkt der DFG. Universität zu Köln. <http://edok01.tib.uni-hannover.de/edoks/e01fn05/504241486.pdf>.
- Tibbo, Helen R. (2012): Placing the Horse before the Cart: Conceptual and Technical Dimensions of Digital Curation. In: Thaller, Manfred (Hrsg.): *Controversies around the Digital Humanities (Historical Social Research – Historische Sozialforschung, HSR; 37/3)*, S. 187–200.
- Zündorf, Irmgard (2007): Historische Fach- und Themenportale. In: Burkhardt, Daniel; Hohls, Rüdiger; Prinz, Claudia (Hrsg.): *Geschichte im Netz: Praxis, Chancen, Visionen. Beiträge der Tagung – hist2006*. Teilband II (*Historisches Forum* 10, II), S. 145–149. Online: http://edoc.hu-berlin.de/histfor/10_II/PDF/HistFor_2007-10-II.pdf.

Geisteswissenschaftliche Aufbereitung von Daten mit TEI

von Werner Wegstein und Wolfgang Pemppe

1 Digitalisierung und Auszeichnung von (Text-) Daten

Auf den ersten Blick öffnet das Thema „Geisteswissenschaftliche Aufbereitung von Daten“ die Sicht auf ein sehr weites Feld. Die Schwelle, die für fast jede geisteswissenschaftliche Aufbereitung von Daten anfangs überschritten werden muss, ist die der Digitalisierung ursprünglich nicht digitaler sprachlicher Quelldaten, bislang vorwiegend Textzeugnisse in allen möglichen medialen Formen. Die technischen Rahmenbedingungen der Digitalisierung für die Welt der Texte sind festgehalten in den Maßstäbe setzenden „DFG-Praxisregeln Digitalisierung“ (DFG 2013), die – gerade erst neu überarbeitet – dem Ziel dienen, „die Planung von Digitalisierungsprojekten zu erleichtern“ und „durch die Formulierung von Standards einen Beitrag zur Nachhaltigkeit und Zukunftsfähigkeit der unterstützten Projekte“ (ebd.: 4) zu leisten. Dementsprechend reicht das thematische Spektrum der Praxisregeln von den technischen und materialspezifischen Parametern der Bilddigitalisierung über die Beschreibung und Erschließung der Digitalisate durch Metadaten in Standardformaten, Grundsätze für die Erstellung digitaler Volltexte bis hin zu Fragen der persistenten Adressierung, der Bereitstellung der Digitalisate für die Öffentlichkeit – Rechtfragen, Open Access und Fragen der Visualisierung (DFG-Viewer) eingeschlossen – sowie Perspektiven der Langzeitverfügbarkeit. In diesem Umfeld ist das Thema unseres Beitrags auf den Bereich der Zeichenkodierung und der generischen Auszeichnung der Volltexte ausgerichtet, wobei die benachbarten Arbeitsbereiche der Digitalisierung hier auch eine Rolle spielen und zu berücksichtigen sind.

Das Akronym TEI im Titel steht für die „Text Encoding Initiative“, gegründet 1987 mit dem Ziel, gemeinschaftlich einen Standard zur Darstellung von Texten in digitaler Form zu entwickeln: “The mission of the Text Encoding Initiative is to develop and maintain a set of high-quality guidelines for the encoding of humanities texts, and to support their use by a wide com-

munity of projects, institutions, and individuals”¹. Die Summe an Arbeit, die die TEI-Community seit mehr als 25 Jahren dafür aufgewandt hat und weiterhin aufwendet, wird repräsentiert durch die vom TEI-Konsortium herausgegebenen *Guidelines for Electronic Text Encoding and Interchange*.² Darin wird ein hardware- und softwareunabhängiges Kodierungskonzept beschrieben, das auf dem Unicode-Standard als Zeichenformat und dem W3C-Standard XML (und damit verwandten Standards) als Auszeichnungssprache aufbaut.

Die Gründe dafür sind: XML ist nicht abhängig von spezifischer Hardware oder Software und grundsätzlich erweiterbar. Es ist auf eine generische Textauszeichnung ausgerichtet, die vorrangig beschreibt, welches Textphänomen an der markierten Stelle vorliegt – und nicht, wie es verarbeitet bzw. dargestellt werden soll. XML unterscheidet dabei zwischen syntaktisch korrekter Kodierung einerseits und einer Validierung der Textkodierung. Die Validierung basiert auf der Vorstellung, dass Texte einem Dokumenttyp (z.B. dem Dokumenttyp Roman, Drama oder Wörterbuch) zugeordnet werden können. Sie erfolgt auf der Basis einer spezifischen Dokumenttypdefinition, in der die Struktureigenschaften des Dokumenttyps – in je nach Projekterfordernissen unterschiedlicher Granularität – festgelegt werden. In der aktuellen Version TEI P5 geschieht das formal über ein Schema (z.B. nach dem Standard-XML-Schema oder nach RelaxNG). Damit lässt sich unmittelbar und direkt prüfen, ob die mittels des Schemas definierten TEI-XML-Strukturen in dem so kodierten Dokument auch definitionsgemäß angewandt sind.

In den grundsätzlich flüchtigen elektronischen Text wird auf diese Weise mit der TEI-Kodierung ein robustes und validierbares Strukturgerüst einge-
zogen, das in vielfältiger Weise nutzbar ist.³ Als Struktureinheiten für den Aufbau dieses Strukturgerüsts dienen sogenannte ‚Elemente‘. Sie bilden eine Hierarchie und können für die Kodierung zusätzlich durch Attribute erweitert werden – Formeigenschaften, die die Verwendung des jeweiligen Elements

1 TEI: Goals and Mission, <http://www.tei-c.org/About/mission.xml>

2 Jede neue Ausgabe der ‚Guidelines‘ wird einmal gedruckt und danach bis zur nächsten Ausgabe permanent in elektronischer Form aktualisiert – aktuelle Druckversion: TEI P5 (2008); aktuelle Internetversion: Version 2.5.0 (last updated on 26th July 2013): <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html> (8.12.2013).

3 dazu ausführlich: TEI P5 (2008) „v. A Gentle Introduction to XML“ (s. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/SG.html>)

an der konkreten Stelle im Text differenzieren oder kennzeichnen. Was Leistung und Bandbreite betrifft, so stehen im TEI-Kodierungssystem über 400 vordefinierte Elemente zur Verfügung, organisiert in 21 Modulen, die im Wesentlichen auf Standardbestandteile von Texten, Dokumenttypen und auf besondere Anwendungsszenarien ausgerichtet sind. Zur feineren Differenzierung und Kennzeichnung können zusätzlich über 200 Attribute eingesetzt werden. Es leuchtet ein, dass sich damit ein breites Spektrum von Datenstrukturen zur Beschreibung der Eigenschaften elektronischer Texte generieren lässt, das der Anwender in den Geisteswissenschaften präzise auf die Anforderungen seines Projekts ausrichten kann. Dass bei komplexen Textphänomenen schnell wortreiche XML-Kodierungskomplexe zusammenkommen, stört dabei nicht, denn sie können auf einfachem Weg bei der Anzeige ausgeblendet oder mittels einer XSLT-Prozedur ohne manuellen Eingriff in ein für die Präsentation optimiertes Anzeigeformat transformiert werden.

Wie die Inhaltsmodelle der TEI-Elemente definiert sind und wie man damit umgeht, wird in den ‚Guidelines‘ im Detail beschrieben. Zur Generierung eines Schemas, gegen das man den kodierten Text schließlich validieren kann, bietet sich zum Beispiel das Programm „ROMA“ auf der TEI-Website an.⁴ Mit der Einstellung „TEI absolute bare“ erzeugt ROMA für Demonstrationszwecke das kleinstmögliche valide TEI-P5-Dokument. Es besteht lediglich aus dem Satz „This is about the shortest TEI document imaginable.“ und soll dazu dienen, die Konstruktion der TEI-Kodierungsstrukturen verständlich zu machen. So sieht es aus:

```
<?xml version="1.0" encoding="UTF-8"?>
<TEI version="5.0" xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>The shortest TEI Document Imaginable</title>
      </titleStmt>
      <publicationStmt>
        <p>First published as part of TEI P2, this is the P5 version using a name
          space.</p>
      </publicationStmt>
      <sourceDesc>
        <p>No source: this is an original work.</p>
```

→

4 Roma: generating customizations for the TEI: <http://www.tei-c.org/Roma/>

```

</sourceDesc>
</fileDesc>
</teiHeader>
<text>
  <body>
    <p>This is about the shortest TEI document imaginable.</p>
  </body>
</text>
</TEI>

```

Nach dem Wurzelement <TEI> folgen 13 Zeilen zur Einbettung der Metadaten im <teiHeader>, die einen ersten Eindruck von den hierarchisch geschachtelten Elementstrukturen aus Elementanfang (z.B. <teiHeader>) und Elementende (entsprechend </teiHeader>) vermitteln. Die Elemente <text> und <body> sind notwendig, damit ein minimaler Absatz, kodiert mit dem Element <p>, mit dem Satz „This is about the shortest TEI document imaginable.“ in eine valide Struktur eingebettet ist. <teiHeader> und <text> sind zwingend erforderliche Rahmenelemente für ein gültiges TEI-Dokument.

2 Bonaventuras ‚Nachtwachen‘ als Praxisbeispiel

Aufwand und Nutzen der TEI-Kodierungspraxis versuchen wir an dem Textanfang des Romans ‚Nachtwachen. Von Bonaventura‘, Penig 1805 (bzw. 1804) zu erläutern. Bonaventura ist ein Pseudonym, die ‚Nachtwachen‘ sind das einzige unter diesem Autornamen überlieferte Werk, das nur von dem Verleger Dienemann in Penig in zwei Formen verbreitet wurde: im ‚Journal von neuen deutschen Original Romanen in 8 Lieferungen jährlich. Dritter Jahrgang 1804. Siebente Lieferung. Nachtwachen. Penig 1804 bey F. Dienemann und Comp.‘ und als Buch unter dem Titel ‚Nachtwachen. Von Bonaventura. Penig 1805, bei F. Dienemann und Comp.‘. Von diesem Buch existiert ein digitales Faksimile des Exemplars an der SUB Göttingen,⁵ das wir als Basis für unser Editionsbeispiel verwenden.

Zur Forschungsgeschichte des Textes sei hier nur so viel festgehalten: Nicht nur in literarischen Zirkeln wurde gerätselt, wer sich hinter ‚Bonaventura‘ verbirgt. Auch in Literatur- und Sprachwissenschaft wurde die Autorfrage kontrovers diskutiert, bis 1973 Jost Schillemeit mit literaturwissenschaftlicher Methodik und Horst Fleig annähernd zeitgleich mit lexikali-

⁵ <http://resolver.sub.uni-goettingen.de/purl?PPN519061209>

schen Analysen den Braunschweiger Theaterdirektor August Klingemann als Autor identifizierten. Endgültig abgeschlossen schien die Autordiskussion, als 1987 Ruth Haag in der Sammlung Diederichs an der Universitätsbibliothek Amsterdam neun Briefe von Klingemann und ein siebenseitiges Manuskript mit Lebens- und Werkdaten von Klingemann fand, in dem von Klingemanns eigener Hand am Rand von S.3 der Nachtrag ergänzt ist „Nachtwachen von Bonaventura, Penig Dienemann 1804“.⁶ Dennoch ist die Autordiskussion nicht verstummt.

Ina Braeuer-Ewers stellt in ihrer Dissertation die Grundsatzfrage, „wie beweiskräftig die sogenannte Sprache der Tatsachen wirklich ist“ und schließt zwar Klingemann nicht als Autor aus, stellt aber fest, dass nichts dafür spricht, dass er „die Nachtwachen alleine geschrieben haben muss“ (Braeuer-Ewers 1994: 13), und verweist darauf, dass Klingemann in seinen Jenaer Studienzeiten mit Brentano befreundet gewesen sei. Ebenso wertet Nicola Kaminski (2001) – mit Bezug auf die Argumentation von Mielke, „dass ein Verfasserbeweis, der sich auf einen ambivalenten Text stützen muss, unmöglich ist“ (ebd.: 43, Anm. 13) – den Wert des Amsterdamer Zeugnisses ab und rückt über eine anagrammatische Umdeutung von „Bonaventura“ zu „Brentano u. v. a.“, wobei das u. v. a. für „und von Arnim“ oder „und viele andere“ stehen könnte, wieder Brentano in den Vordergrund. Jedenfalls wird die Rückbesinnung auf den Text als letztendlich entscheidende Instanz für alle Beweisfragen in den jüngeren Arbeiten deutlich. Die Anforderungen an die elektronische Kodierung des Textes sind darauf auszurichten.

3 Der Weg zum elektronischen Referenztext

3.1 Leitprinzipien am Modellfall „Erste Nachtwache.“

Die bisherigen Ausgaben von Michel (1904) bis Schillemeit (2012)⁷ waren angelegt auf eine mehr oder weniger diplomatische Edition des Textes: „Wortlaut, Orthographie und Interpunktion sind die des Erstdrucks (Nachtwachen. Von Bonaventura. Penig 1805 [1804] bey F Dienemann und Comp.). Lediglich offensichtliche Druckfehler wie ‚Darwie‘ für ‚Darwin‘ und Ähnliches wurden verbessert, außerdem die drucktechnisch bedingten

6 zitiert nach Schillemeit (2012: 264) (Nachbemerkung von Rosemarie Schillemeit)

7 weitere Ausgaben: Frank (1912), Steinert (1914)

Schreibungen Ae, Oe, Ue des Originals durch die heute üblichen ersetzt“ (Schillemeit 1974: 201). Zur editionsphilologischen Diskussion darüber, was ein Fehler ist, soll hier der Verweis auf Zeller (1971) und Plachta (1997) genügen. Ferner ist festzuhalten, dass auch schon die Edition von Schillemeit (1974) nicht fehlerfrei war⁸ und auch in der Neuedition (2012) z. B. an prominenter Stelle neu in den Text eingegriffen wurde: Der allerletzte Satz des Textes „Und der Wiederhall im Gebeinhaus ruft zum letztenmale — Nichts! —“⁹ ist durch die kommentarlose Ergänzung eines Abführungszeichens am Textende in einer Weise verändert, die bis auf die Interpretation durchschlagen könnte (vgl. Schillemeit 2012: 136).¹⁰

Eine elektronische Kodierung des Textes nach TEI-Richtlinien bietet jetzt die Gelegenheit, den Druck – so wie er ist, mit allen „Fehlern“ – als Basistext zu erstellen, alle Veränderungen an dem Text kenntlich zu machen und Verfahren zu nutzen, die ohne manuellen Eingriff Lesetexte mit Korrekturen oder orthografischen Normierungen erzeugen, sodass den an dem Text Interessierten die Wahl überlassen bleibt, in welcher Gestalt sie die ‚Nachtwachen‘ lesen wollen. In jedem Fall bleibt die Verbindung zum digitalisierten Faksimile, sodass auf jeder Seite an einem Originaldruck überprüft werden kann, was wirklich geschrieben steht.

Ausgangspunkt für die Edition ist eine typografische Analyse, die alle Merkmale ermittelt, die in der elektronischen Kodierung festgehalten werden sollen. Soweit Standards zur Kennzeichnung in Gebrauch sind, wird auf diese Bezug genommen. So richtet sich die Differenzierung der Formkriterien nach dem W3C-Standard Cascading Style Sheets.¹¹

3.2 *Die Digitalisate*

Die Digitalisierung der Erstausgabe¹² von 1804/05 erfolgte 2006 seitens des Digitalisierungszentrums der Niedersächsischen Staats- und Universitäts-

8 Vgl. z.B. Kaminski (2001), S. 45, Anm. 24, bzw. S. 77, Anm. 130.

9 <http://resolver.sub.uni-goettingen.de/purl?PPN519061209>; Text S. 298, Originalseite 296

10 Vgl dazu Kaminski (2001), S. 105, Anm. 196.

11 <http://www.w3.org/TR/CSS2/>

12 <http://gdz.sub.uni-goettingen.de/dms/load/mod/?PPN=PPN519061209>

bibliothek Göttingen (GDZ) im Rahmen des DigiWunschbuch-Dienstes¹³. Bei der Digitalisierung im GDZ wird neben den eigentlichen Scans in unterschiedlichen Auflösungen und Größen auch ein METS-File generiert, das einerseits als Container für die bibliografischen und technischen Metadaten dient und andererseits die Grafikdateien in jeweils einer logischen (inhaltliche Kriterien, z.B. Kapitel) und physischen Struktur (Seitenabfolge) miteinander verknüpft. Das hierbei verwendete Metadatenprofil¹⁴ entspricht den Vorgaben zur Nutzung des sogenannten DFG-Viewers¹⁵, einem Online-Tool zur Visualisierung digitalisierter Werke aus Bibliotheksrepositorien. Zur Darstellung der Nachtwachen siehe DFG-Viewer (2013).

Eine synoptische oder wie auch immer geartete Verknüpfung der Seitenansicht mit den zugehörigen Textdaten wird vom DFG-Viewer bislang nicht unterstützt. Für die Umsetzung einer solchen Lösung kann entweder auf ein bereits bestehendes Framework wie SADE¹⁶ oder eSciDoc¹⁷ bzw. ViRR¹⁸ (Virtueller Raum Reichsrecht) zurückgegriffen oder mittels geeigneter Transformationen (XSLT) eine eigene, leichtgewichtige HTML-basierte Lösung implementiert werden.

Hierzu muss zunächst die Verknüpfung zu den Bilddaten im digitalen Volltext verankert werden:

```
<facsimile>
<graphic xml:id="FILE_0001_PRESENTATION"
  url="http://gdz.sub.uni-goettingen.de/tiff/PPN519061209/00000001.tif"/>
<graphic xml:id="FILE_0002_PRESENTATION"
  url="http://gdz.sub.uni-goettingen.de/tiff/PPN519061209/00000002.tif"/>
<graphic xml:id="FILE_0003_PRESENTATION"
  url="http://gdz.sub.uni-goettingen.de/tiff/PPN519061209/00000003.tif"/>
<!-- usw. -->
<graphic xml:id="FILE_0300_PRESENTATION"
  url="http://gdz.sub.uni-goettingen.de/tiff/PPN519061209/00000300.tif"/>
</facsimile>
```

13 <http://www.digiwunschbuch.de>

14 <http://dfg-viewer.de/profil-der-metadaten/>

15 <http://dfg-viewer.de/ueber-das-projekt/>

16 Scalable Architecture for Digital Editions, <http://www.bbaw.de/telota/software/sade>

17 <https://www.escidoc.org>

18 <http://virr.mpg.de> – Unterstützung für Volltexte geplant

Der facsimile-Block wurde mithilfe eines XSLT-Stylesheets aus den vom GDZ erstellten METS-Daten generiert, wobei jeweils URL und METS File-Id der Scans mit der höchsten Auflösung (TIFF) als Referenz übernommen wurden. Die File-Id wird dann in den jeweiligen <pb>-Elementen referenziert:

```
<pb n="002" facs="#FILE_0004_PRESENTATION"/>
```

Es wäre natürlich auch möglich gewesen, die URL des jeweiligen Seiten-Scans direkt in @facs anzugeben, hierauf wurde jedoch u.a. im Sinne einer besseren Lesbarkeit der Textdaten verzichtet. Sollten die Digitalisate jemals in einem Forschungsdaten-Repository liegen, genügt es, den facsimile-Block einfach auszutauschen bzw. entsprechend zu modifizieren. Der Import ins TextGrid-Repository war zur Entstehungszeit dieses Beitrags aus technischen Gründen nicht möglich. Gleichwohl existiert mit koLibRI¹⁹ ein mächtiges Tool u.a. für den Batch-Import nach TextGrid, das auch den Ingest von Kollektionen ermöglicht, die über ein DFG-Viewer-METS-File modelliert werden.²⁰

3.3 Die Titelei: <teiHeader>

Bei der Auszeichnung und Zusammenstellung der für die Titelei relevanten Daten wurde weitestgehend den Empfehlungen für die TextGrid-Kernkodierung (vgl. Blümm/Wegstein 2008) entsprochen. Der teiHeader besteht demnach aus den drei Blöcken fileDesc, encodingDesc und revisionDesc, wobei fileDesc die tiefste Binnenstruktur aufweist und gemäß TEI-Standard als Einziges dieser Elemente obligatorisch ist. Es besteht in unserem Fall aus:

- titleStmt, der Titelei im engeren Sinne,
- publicationStmt, Informationen zur Verfügbarkeit des Textes / der Daten,
- notesStmt, ergänzenden Angaben, sowie
- sourceDesc, dem Container für bibliografische Metadaten.

In titleStmt muss zusätzlich zum Titel der elektronischen Transkription (in <title>) der Herkunft der verwendeten bzw. verknüpften Ressourcen Rechnung getragen werden:

19 kopal Library for Retrieval and Ingest, http://kopal.langzeitarchivierung.de/index_koLibRI.php

20 <https://dev2.dariah.eu/wiki/pages/viewpage.action?pageId=6783190>

```

<respStmt>
  <resp>Digitalisierung</resp>
  <name>Digitalisierungszentrum der Niedersächsischen Staats- und
    Universitätsbibliothek Göttingen</name>
</respStmt>
<respStmt>
  <resp>Transkription und Textauszeichnung</resp>
  <name>Werner Wegstein</name>
  <name>Wolfgang Pempe</name>
</respStmt>

```

während in `publicationStmt` sowohl die bereits im Rahmen der Digitalisierung erfolgte Vergabe von Identifiern (`idno`) als auch etwaige Nutzungs- und/oder Lizenzbestimmungen (`availability`) dokumentiert sind:

```

<publicationStmt>
  <authority>Niedersächsische Staats- und Universitätsbibliothek Göttingen,
    Digi talisierungszentrum</authority>
  <idno type="mods:recordIdentifier">PPN519061209</idno>
  <idno type="purl">http://resolver.sub.uni-
    goettingen.de/purl?PPN519061209</idno>
  <idno type="gbv-ppn">PPN519061209</idno>
  <idno type="PPNanalog">PPN133911934</idno>
  <idno type="ADBID">300247</idno>
  <availability>
    <p>Digitalisate: Public Domain (<ref target="http://creativecommons.org/pub-
      lic domain/zero/1.0/legalcode.txt">CC021</ref>), Transkription und XML-
      Daten: <ref target="http://creativecommons.org/licenses/by-sa/3.0/legal-
        code">CC BY-SA 3.0</ref>
    </p>
  </availability>
</publicationStmt>

```

Auf die explizite Codierung der Nutzungsbestimmungen als licence wurde im Sinne einer besseren Lesbarkeit verzichtet.

Zusätzlich Bemerkungen können in `notesStmt` untergebracht werden:

21 Zur Zeit der Abfassung dieses Beitrags noch nicht endgültig geklärt – diese Angabe stellt insofern nur ein Beispiel dar und ist in keiner Weise autoritativ.

```
<notesStmt>
  <note>Transkription und Textauszeichnung unter Berücksichtigung der
    Magis terarbeit von Stephanie Rill (Würzburg 2007)</note>
</notesStmt>22
```

Außerhalb des oben genannten Kanons folgen nach fileDesc noch Angaben zur Auflösung der für Formatierungszwecke verwendeten Kürzel sowie die Deklaration des zur Darstellung der Textdaten vorgesehenen Standards (styleDefDecl):

```
<styleDefDecl scheme="css"/>
<tagsDecl>
  <rendition xml:id="cnt" scheme="css">text-align: center;
</rendition>
  <rendition xml:id="mt-1" scheme="css">margin-top: 1em;
</rendition>
  <rendition xml:id="mt-7" scheme="css">margin-top: 7em;
</rendition>
  <rendition xml:id="ti" scheme="css">text-indent: 1em;</rendition>
  <rendition xml:id="flt" scheme="css" scope="first-letter">font-size: 16pt;
</rendition>
</tagsDecl>
```

3.4 *Typografische Analyse und Textauszeichnung*

Schrift

Im Druck sind drei Schriften verwendet: für die deutschen Textpartien die Frakturvariante der lateinischen Schrift (ISO 15294: Latf), für lateinische und englische Fremdwörter die Antiqua der lateinischen Schrift (ISO 15294: Latn), allerdings nicht konsequent. Hinzukommen drei Buchstaben aus der griechischen Schrift (ISO 15294: Grek). An Textauszeichnungen erscheinen frakturschriftbedingt ‚gesperrt‘ (css: letter-spacing) für Überschriften und Hervorhebungen im Text sowie Initialen am Kapitelanfang (css: first-letter).

²² Die Autoren dieses Beitrags danken Stephanie Rill für die Zustimmung, ihre Magisterarbeit für die Edition des Originaldrucks nutzen zu dürfen.

Erste Nachtwa che,

Die Nachtstunde schlug; ich hüllte mich in meine abenteuerliche Vermummung, nahm die Pike und das Horn zur Hand, ging in die Finsterniß hinaus und rief die Stunde ab, nachdem ich mich durch ein Kreuz gegen die bösen Geister geschützt hatte.

Es war eine von jenen unheimlichen Nächten, wo Licht und Finsterniß schnell und seltsam mit einander abwechselten. Am Himmel flogen die Wolken, vom Winde getrieben, wie wunderliche Riesenbilder vorüber, und der Mond erschien und verschwand im raschen

I

Seitengliederung

Sofern kein neues Kapitel beginnt, steht die Seitenzahl oben außen, danach folgen maximal 22 Zeilen. Beginnt auf einer Seite ein neues Kapitel, ist keine Seitenzahl gesetzt und der Text beginnt nach sechs Leerzeilen mit der Kapitelüberschrift, gesperrt und zentriert. Nach einer Leerzeile folgt ein Trenn-

strich (Grafik hr1.png), zentriert. Nach einer weiteren Leerzeile beginnt das Kapitel stumpf mit einer zweistufigen Initiale. Alle 16 Seiten ist am unteren Seitenrand rechts eine fortlaufende Zahl als Bogensignatur gedruckt. Das Kapitelende ist ebenfalls mit einem (etwas stärkeren Trennstrich markiert (Grafik hr2.png). Anmerkungen werden im Text durch einen hochgestellten Asterisk und eine runde Klammer markiert und – mit einer Zeile Abstand – zusammen mit dem Verweiszeichen am unteren Seitenende in kleinerer Schrift gesetzt.

Absatzgliederung

Folgende Absatzformen sind im Text vertreten: Kapitelanfang stumpf mit Leerzeile (css: margin-top) und Initiale (css: first-letter), Absatzmarkierung durch Einrückung (css: text-indent), Absatzmarkierung durch Leerzeile (css: margin-top) und Einrückung (css: text-indent), sowie Zwischenüberschriften (wie Kapitelüberschriften gesperrt) zur Untergliederung der Kapitel und Fußnoten zum Text am Seitenende. In der 10. Nachtwache (S. 178) und in der 14. Nachtwache (S. 238) erscheint zusätzlich eine Trennlinie zwischen zwei Absätzen. An einigen Stellen ist – für die Zeit nicht untypisch – eine syntaktische Konstruktion verwendet, die in einem Absatz beginnt und eine Phrase enthält, die typografisch wie eine Zwischenüberschrift gesetzt ist, mit entsprechenden Leerzeilen zur Kennzeichnung als Überschrift, und die im nachfolgenden Absatz weitergeführt wird.

Orthografie und Interpunktion

Mit Ausnahme der Ligaturen (z.B. ch und tz) transkribieren wir jeden einzelnen gedruckten Buchstaben, auch die frakturbedingten Großbuchstaben Ae, Oe und Ue anstelle von Ä, Ö und Ü in der Gegenwartssprache. Die Differenzierung des Fraktur-s in Lang-s (Schaft-s) und Rund-s (Schluss-s) haben wir nicht übernommen. Ebenso bilden wir die Umlaute der Kleinbuchstaben ä, ö und ü – in der Frakturschrift mit übergeschriebenem kleinen e – mit den heute üblichen Buchstabenformen ab.

Zur Kodierung der Interpunktion stellen die TEI Guidelines das Element <pc> (punctuation character) mit entsprechenden Attributen für eine feinere Differenzierung bereit. Wir beschränken uns hier, der Übersichtlichkeit halber, auf die Kodierung von zwei Phänomenen: die Worttrennung am Zeilenende und die Verwendung von Anführungszeichen. Die Worttrennung am Zeilenende haben wir belassen und markiert, auch bei Worttrennungen in k-k bzw. z-z in Wörtern, die im Text mit der Schreibung -ck- bzw. -tz- erschei-

nen. Die für die deutsche Sprache kennzeichnenden Funktionsunterschiede von Trennungsstrich/Trennsignal, Bindestrich und Ergänzungsbindestrich²³ halten wir mit den Kürzeln "Ts", "Bs" bzw. "Es" im Attribut „function“ des Elements <pc> fest. In ähnlicher Weise kodieren wir die Verwendung von Anführungszeichen.

Die TEI-Kodierung der ersten Textseite der ‚Nachtwachen‘ sieht damit folgendermaßen aus:

```
<body>
<div type="chapter" n="N01">
  <pb n="001" facs="#FILE_0003_PRESENTATION"/>
  <lb n="01"/><head rendition="#mt-7 #lsp-5">Erste Nachtwache.</head>
  <lb n="02"/><figure>
    <graphic url="hr1.png" rendition="#mt-1 #cnt"></graphic>
  </figure>
  <lb n="03"/><p n="N01.01" rendition="#mt-1 #flt">
    <s n="N01.01.01">Die Nachtstunde schlug; ich hüllte mich in
  <lb n="04"/>meine abenteuerliche Vermummung, nahm
  <lb n="05"/>die Pike und das Horn zur Hand, ging in die
  <lb n="06"/>Finsterniß hinaus und rief die Stunde ab,
  <lb n="07"/>nachdem ich mich durch ein Kreuz gegen die
  <lb n="08"/>bösen Geister geschützt hatte.</s></p>
  <lb n="09"/><p n="N01.02" rendition="#ti">
    <s n="N01.02.01">Es war eine von jenen unheimlichen Näch
    <pc function="Ts">-</pc>
  <lb n="10" break="no"/>ten, wo Licht und Finsterniß schnell und selt
    <pc function="Ts">-</pc>
  <lb n="11" break="no"/>sam mit einander abwechselten.</s>
    <s n="N01.02.02">Am Himmel
  <lb n="12"/>flogen die Wolken, vom Winde getrieben, wie
  <lb n="13"/>wunderliche Riesenbilder vorüber, und der
  <lb n="14"/>Mond erschien und verschwand im raschen
  <fw type="sig" place="bottom">1</fw>
```

Diese Kodierung soll die Formeigenschaften der zugehörigen gedruckten Textseite angemessen beschreiben.

²³ Zur Bindestrich-Problematik im Detail siehe Gallmann (1985: 83–101, 219–222).

Die linearen Strukturen des Textes, Seite und Zeile, werden in den Elementen <pb> (page break) und <lb> (line break) festgehalten. Mit dem Attributwert von n (für ‚number‘ oder ein anderes Kennzeichen) innerhalb der Elemente wird die Seitenzahl bzw. die Zeilennummer kodiert. Das Attribut @facs in <pb> stellt die Beziehung zu dem digitalisierten Bild her. Die Textstrukturen beginnen mit <div> zur Kennzeichnung der Einheit „Erste Nachtwache.“, festgehalten im Attribut @type mit dem Wert "chapter". Das Attribut n wird hier und in den anderen textbezogenen Elementen zum Aufbau eines seiten- und zeilenunabhängigen Referenzsystems genutzt. @n="N01" ist das Kennzeichen für die erste Nachtwache. Das Element <head> markiert „Erste Nachtwache.“ als Überschrift, das @rendition-Attribut hält die Formeigenschaften fest, "mt-7" steht für margin-top = 7 em, d.h. voraus gehen sieben Leerzeilen, "lsp-5" steht für „letter-spacing“ – also gesperrt – mit dem Wert 0,5 em und </head> markiert das Ende der Überschrift.

Die Auflösung der Formatierungskürzel erfolgt im <teiHeader>, siehe oben.

Der Text beginnt nach dem line break in Zeile 3 mit dem Element <p>, das Absätze markiert. Das @n-Attribut kennzeichnet diesen Absatz mit dem Referenzwert "N01.01", d.h. das ist der erste Absatz in der ersten Nachtwache. Das Attribut @rendition hält die Formeigenschaften dieses Absatzes fest: "mt-1" steht wieder für margin-top = 1, d.h. eine vorausgehende Leerzeile, "flt" steht für „first-letter“ und drückt aus, dass der erste Buchstabe des Absatzes als Initiale gestaltet ist. Es folgt das Element <s> für „s-unit“ eine satzartige Gliederungseinheit des Textes; auch hier dient das @n-Attribut "N01.01.01" als Referenzwert mit der Bedeutung: „Erste Nachtwache.“, erster Absatz, erster Satz! Von da läuft der Text fort über die Zeilenmarkierungen bis zum Elementende mit </s> und dem Absatzende mit </p>, d.h. der erste Absatz besteht nur aus einem Satz.

Es folgt der zweite Absatz, eingerückt (rendition="#ti" für „text-indent“), aber ohne Leerzeile. Mit <s n="N01.02.01"> folgt der erste Satz des zweiten Absatzes. Interessant ist hier das Zeilenende der 9. Zeile: auf den Wortanfang „Näch“ folgt ein <pc>-Element mit dem Attribut @function=(Trennungsstrich/Trennsignal), d.h. eine Markierung dafür, dass das Wort nicht am Zeilenende endet. Auch der line break zu Beginn der 10. Zeile mit dem Wert @break="no" hält fest, dass der Zeilenanfang nicht mit einem Wort beginnt. Mit dieser Kennzeichnung kann später die Worttrennung am Zeilenende ohne manuellen Eingriff durch eine Prozedur korrekt aufgelöst werden.

Bleibt schließlich noch das letzte Element <fw> am Seitenende. Es hält den Wert der Bogensignatur auf der ersten Seite fest.

Eine direkte Rede auf S. 3 verdient noch besondere Aufmerksamkeit. Der Bildausschnitt ab Zeile 5 zeigt folgendes Satzphänomen, das in der Zeit noch gebräuchlich ist,²⁴ das sich in den ‚Nachtwachen‘ aber nur an dieser Stelle noch findet:

„O du, der du da oben dich herumtreibst,
 „ich verstehe dich wohl, denn ich war einst
 „deinesgleichen! Aber ich habe diese Beschäf-
 „tigung angesetzt gegen ein ehrliches Hand-
 „werk, das seinen Mann ernährt, und das
 „für denjenigen, der sie darin aufzufinden
 „weiß, doch keinesweges ganz ohne Poesie ist.“

```
<lb n="05"/><p n="N01.07" rendition="#mt-1 #i">
  <s n="N01.07.01"><pc function="qmOpen">.,</pc>O du, der du
    da oben dich herumtreibst,
  <lb n="06"/><pc function="qmCont">.,</pc>ich verstehe dich wohl,
    denn ich war einst
  <lb n="07"/><pc function="qmCont">.,</pc>deinesgleichen!</s>
  <s n="N01.07.02">Aber ich habe diese Beschäf<pc function="Ts">-</pc>
  <lb n="08" break="no"/><pc function="qmCont">.,</pc>tigung
    aufgegeben gegen ein ehrliches Hand<pc function="Ts">-</pc>
  <lb n="09" break="no"/><pc function="qmCont">.,</pc>werk, das
    seinen Mann ernährt, und das
  <lb n="10"/><pc function="qmCont">.,</pc>für denjenigen, der sie
    darin aufzufinden
  <lb n="11"/><pc function="qmCont">.,</pc>weiß, doch keinesweges ganz ohne
    Poesie ist.</s>
```

24 Vgl. Parkes (1992), Glossary S. 303, s.v. diiple (Plates 36, 38, 40 und 55-6).

Das öffnende Anführungszeichen zur Kennzeichnung direkter Rede wird hier nicht nur am Anfang der Rede gesetzt, sondern bis zum schließenden Anführungszeichen vor jeder Zeile wiederholt, in der die Rede noch fortgesetzt wird. Dadurch weist die Rede nicht nur mehr als ein öffnendes Anführungszeichen auf, durch die Worttrennung am Zeilenende rückt das Anführungszeichen auch noch in das Wortinnere des getrennten Wortes, wo es eigentlich nicht zu erwarten ist. Die feinere Differenzierung der Kodierung von Anführungszeichen mit dem Elements `<pc>` durch das Attribut „function“ erlaubt es, zwischen öffnenden, wiederholten, schließenden und (in den ‚Nachtwachen‘ gelegentlich) fehlenden schließenden Anführungszeichen zu unterscheiden. Zur Vereinheitlichung der Auszeichnungssprache verwenden wir die Attributwerte gern in englischer Sprache (opening, repeated, closing and missing quotation mark) und der besseren Übersichtlichkeit wegen abgekürzt in ‚Camelcase‘-Form („qmOpen“, „qmRep“, „qmClose“ und „qmMiss“). Diese Art der Kodierung, die bei Bedarf noch weiter verfeinert werden kann, ermöglicht sowohl präzise Angaben über die Interpunktion des Texts und darüber hinaus zusätzlich eine automatische Korrektur in eine vereinfachte bzw. vereinheitlichte Form, falls nötig.

Ein letzter Blick auf den Text soll noch einen weiteren Anwendungsbereich der Kodierung verdeutlichen:

Glaubens und der Poesie. Sie ist die doppelte Beleuchtung in der Corregios Nacht, und verschmilzt den irdischen und himmlischen Strahl zu Einem wunderbaren Glanze.

```
<lb n="11"/>Glaubens und der Poesie.</s><s n="N01.15.04">Sie ist die dop
  <pc function="Ts">-</c>
<lb n="12" break="no"/>pelte Beleuchtung in der <persName nymRef="Corregio,
  Antonio Allegri da" ref="http://d-nb.info/gnd/16012154X">Corregios
  </persName> <name type="painting" ref="http://de.wikipedia.org/wiki/
  Die_Heilige_Nacht_%28Correggio%29">Nacht</name>, und
<lb n="13"/>verschmilzt den irdischen und himmlischen
<lb n="14"/>Strahl zu <hi rendition="#cap">Einem</hi> wunderbaren
  Glanze.</s></p>
```

Auf S. 7, Z. 12, geht es um die ‚Corregios Nacht‘. Die Kodierung von ‚Corregio‘ und von ‚Nacht‘ weicht von den üblichen Wörtern ab, denn bei Corregio handelt es sich um einen Namen, kein Begriffswort, was korrekterweise auch entsprechend kodiert werden sollte. Hilfreich ist dabei die Möglichkeit, nicht nur die korrekte Form des Namens im Attribut @nymRef festzuhalten, sondern zusätzlich mittels des Attributs @ref="http://d-nb.info/gnd/16012154X" einen Link zum Personenkennzeichen der Gemeinsamen Normdatei an der Deutschen Nationalbibliothek herzustellen und damit dafür zu sorgen, dass Personen eindeutig zugeordnet werden. Gleiches gilt auch für ‚Nacht‘, das hier ebenfalls nicht für das Begriffswort ‚Nacht‘ steht, sondern ein Gemälde bezeichnet. Darauf verweist das Attribut @type im Element <name> und über das Attribut @ref wird die Verbindung zu dem entsprechenden Beitrag in ‚Wikipedia‘ hergestellt.

3.5 Ausblick: Perspektiven der Texterschließung

Die im Beitrag vorgeschlagene Aufbereitung der Textüberlieferung mit dem Ziel einer Referenzedition sowohl für die philologische Forschung wie für die am kulturellen Erbe interessierte Öffentlichkeit kombiniert die elektronische Textkodierung auf der Basis der TEI P5 Guidelines mit einem elektronischen Faksimile des Originaldrucks der ‚Nachtwachen. Von Bonaventura‘. Die TEI-Kodierung ist – mit Ausnahme der ‚Corregios Nacht‘ – zunächst nur bis zur Satzebene skizziert, kann aber jederzeit ohne Veränderung der vorhandenen Strukturen bis zur Wortebene und damit zur Lemmatisierung oder – in noch feinerer Auflösung – bis zur Morphemebene und damit zu Wortbildungs- und zu Wortfamilienanalysen verfeinert werden. Auf all diesen Schichten kann das von Seitenbruch und Zeilenfall des Originals unabhängige Referenzsystem genutzt werden, um eine differenzierte Analyse zu jeder einzelnen Nachtwache zu erstellen, als Basis für die Aufarbeitung der literaturwissenschaftlichen wie der sprachwissenschaftlichen Diskussion²⁵ am und im Text. Der Modellfall ‚Corregios Nacht‘ kann dabei vielleicht eine Vorstellung davon vermitteln, wie Motivzitate anderer Autoren in den ‚Nachtwachen‘ und Querbeziehungen zu anderen Texten in der Referenzedition verankert werden können.

25 Vgl. z. B. die Forschungsrichtung von Mielke (1984), Braeuer-Ewers (1995), Böning (1996), Kaminski (2001) bzw. Arnold (2005).

Eine am Text verankerte, EDV-gestützte Aufarbeitung der breitgefächerten Forschungsdiskussion über die ‚Nachtwachen‘ und ihre(n) Autor(en) bzw. Autorin(nen) könnte die Grundlage bilden für eine Abwägung der Interpretationsmöglichkeiten und eine Analyse der philologischen Erkenntnisgrenzen, die die ‚Nachtwachen. Von Bonaventura‘ der philologischen Forschung und damit auch der Autoridentifikation ziehen. Und schließlich hat uns die Beschäftigung mit dem Text der ‚Nachtwachen. Von Bonaventura‘²⁶ und seinen Interpretationen spät noch einmal erinnert an die Studie von Morris Weitz *Hamlet and the Philosophy of Literary Criticism* (Weitz 1964).

Literaturverzeichnis

- Arnold, Barbara (2005): *Lexikographische Studien zu August Klingemann*. Thesis (Ph.D.), Department of German. University of Exeter (GB).
- Blümm, Mirjam; Wegstein, Werner (2008): The TEI header for texts in Baseline Encoding. In: Blümm, Mirjam et al. (Hrsg.): *TextGrid's Baseline Encoding for Text Data in TEI P5 (2007–2009)*, S. 19–27. <https://dev2.dariah.eu/wiki/download/attachments/7439482/baseline-all-en.pdf>.
- Böning, Thomas (1996): *Widersprüche. Zu den „Nachtwachen. Von Bonaventura“ und zur Theoriedebatte* (Rombach Wissenschaft: Reihe Litterae; Bd. 38). Freiburg im Breisgau: Rombach.
- Braeuer-Ewers, Ina (1994): *Züge des Grotesken in den Nachtwachen von Bonaventura*. Berlin, Freie Univ., Diss. (Zugl.: Paderborn u.a.: Schöningh, 1995).
- DFG (2013): DFG-Praxisregeln ‚Digitalisierung‘. http://www.dfg.de/formulare/12_151/12_151_de.pdf.
- Frank, Erich (Hrsg.) (1912): *Clemens Brentano. Nachtwachen von Bonaventura*. Heidelberg: Winters Universitätsbuchhandlung.
- Gallmann, Peter (1985): *Graphische Elemente der geschriebenen Sprache. Grundlagen für eine Reform der Orthographie* (Reihe Germanistische Linguistik; Bd. 60). Tübingen: Niemeyer.
- Kaminski, Nicola (2001): *Kreuz-Gänge. Romanexperimente der deutschen Romantik* (Zugl.: Paderborn u.a.: Schöningh 2001)

26 Die Autoren beabsichtigen, das im vorliegenden Beitrag vorgestellte Projekt als Blog weiterzuführen und ausführlich zu dokumentieren. Die Blog-Adresse lautet: www.bonaventura-blog.de.

- Michel, Hermann (Hrsg.) (1904): *Nachtwachen. Von Bonaventura* (Deutsche Literaturdenkmale des 18. und 19. Jahrhunderts; Bd. 133). Berlin: B. Behr.
- Mielke, Andreas (1984): *Zeitgenosse Bonaventura* (Stuttgarter Arbeiten zur Germanistik; Bd. 132). Stuttgart: Heinz.
- Parkes, Malcolm Beckwith (1992): *Pause and Effect. An Introduction to the History of Punctuation in the West*. Aldershot: Scolar.
- Plachta, Bodo (1997): *Editionswissenschaft. Eine Einführung in Methode und Praxis der Edition neuerer Texte*. Stuttgart: Reclam.
- Rill, Stephanie (2007): *Nachtwachen. Von Bonaventura: Editionsphilologische Erschließung*. Magisterarbeit, Würzburg.
- Schillemeit, Jost (1974): *Nachtwachen / von Bonaventura, August Klingemann*. Mit Ill. von Lovis Corinth. Hrsg. u. mit e. Nachw. versehen von Jost Schillemeit; Frankfurt a.M.: Insel.
- Schillemeit, Jost (2012): *Nachtwachen von Bonaventura : Freimüthigkeiten / August Klingemann*. Hrsg. und komm. von Jost Schillemeit. Göttingen: Wallstein.
- Steinert, Raimund (Hrsg.) (1914): *Nachtwachen. Von Bonaventura*. Nach Rachel Varnhagens Exemplar mit einem Nachwort hrsg. Weimar: Kiepenheuer.
- TEI P5 Guidelines (2008): Guidelines for Electronic Text Encoding and Interchange, by the TEI Consortium. Originally edited by C. M. Sperberg-McQueen and Lou Burnard for the ACH-ALLC-ACL Text Encoding Initiative. Now entirely revised and expanded under the supervision of the Technical Council of the TEI Consortium. Edited by Lou Burnard and Syd Bauman 1.1.0. Last updated on 4th July 2008. Oxford, Providence, Charlottesville, Nancy.
- Weitz, Morris (1964): *Hamlet and the Philosophy of Literary Criticism*. Chicago/London: Faber and Faber.
- Zeller, Hans (1971): Befund und Deutung. Interpretation und Dokumentation als Ziel und Methode der Edition. In: Gunter Martens; Hans Zeller (Hrsg.): *Texte und Varianten. Probleme ihrer Edition und Interpretation*. München: Beck, S. 45–89.

From Fragments to an Integrated European Holocaust Research Infrastructure

Tobias Blanke, Veerle Vanden Daelen, Michal Frankl, Conny Kristel, Kepa J. Rodriguez & Reto Speck

In order to keep remembering why it was judged so important to build a new Europe ‘out of the crematoria of Auschwitz’, the ‘vital link’ between Europe’s past and Europe’s present should, according to the British historian Tony Judt, be taught over and over again (cf. Judt 2005: 830 f.). Every generation of European historians should re-interpret the message of the Second World War and of the Holocaust anew. To be able to discuss, negotiate and teach Europe’s past, the historical research needs to become truly transnational and transcend national borders. In order to achieve this and to enable new, innovative forms of Holocaust research, the European Holocaust Research Infrastructure (EHRI) has been set up by the European Union to create a sustainable complex of services for researchers. EHRI will bring together information about dispersed collections, based on currently more than 20 partner organisations in 13 countries and many other archives. EHRI, which brings together historians, archivists and specialists in digital humanities, strives to develop innovative on-line tools for finding, researching and sharing knowledge about the Holocaust. While connecting information about Holocaust collections, it strives to create tools and approaches applicable to other digital archival projects.

Trans-national research into the Holocaust can be a very challenging undertaking. Holocaust studies rely more than other fields of research on a huge variety of archives. Holocaust archives are fragmented and dispersed all over the world, making access complicated, if not impossible and very time-consuming. Researchers therefore have to deal with different archival systems, laws, rules on access and copyright, cataloguing systems, or in short: different archival cultures.

The fragmentation of archival sources does not only result from the wide geographic scope of the Holocaust, but also from the Nazi attempts to destroy the evidence. Moreover, and closely related to the geographical scope, Holocaust sources were written in many different languages, the language of the documents not necessarily being the same as the cataloguing and descrip-

tion language, thus further complicating the picture. Furthermore, Holocaust survivors migrated to places across the world taking documentation with them, and a plethora of documentation projects was developed after the Second World War.

Next to these challenges some new ones have emerged recently. In the past decades more and more specific collections have been set up, in many regional centres on research and commemoration. The opening up of archives in Eastern Europe, and in particular in Eastern Germany, and the opening of formerly classified archives in Western Europe has resulted in a substantial increase in available source material. At the same time, the number of institutions in European countries that hold Holocaust-related collections and are active in the field of research and commemoration, has increased since 1989 (especially in Eastern Europe but also in Germany and most other European states). These institutions, old and new, have their own collections and their own (increasingly digital) archival infrastructures, which often do not support scholarly requirements. Different institutions use their own distinct systems and different metadata schemas. Many different languages are used in the original documents as well as in catalogues, necessitating translation and hampering comparability.

Until 1989 the United States, Israel and Western Europe were the main centers for Holocaust research. Auschwitz became the symbol for the Holocaust worldwide, because it was the largest death camp, but also because it was the camp where Jews from Western and Central Europe were murdered. However, the vast majority of Holocaust victims lived and was murdered in Eastern Europe. Research and documentation on this part of Europe is still far more difficult to conduct than on Germany or Western Europe.

Finally, one of the major challenges for every scholar of the Holocaust is to deal appropriately with the prevalence of perpetrators' sources to avoid muting the voices of the persecuted Jews. The documents of Jewish organizations or relief organizations often followed the fate of their owners; they were in many cases destroyed or dispersed. For instance, to gain insight into the activity of the Jewish refugee organisations in Prague in the 1930s, a researcher has to study the fragments of reports saved in several archives, especially in the USA, Israel, Czech Republic and Germany.¹ And while there are numerous testimonies given by Holocaust survivors after the liberation, original diaries, letters and/or testimonies from the time of persecution are

¹ See the information about sources in Čapková/Frankl (2012: 17 f.).

more difficult to find. Over the last years a growing consensus has emerged in Holocaust historiography that Jewish sources and views have to be more integrated into the narrative(s) of the Holocaust.²

By connecting collections from different archives and countries with research, EHRI can make a significant contribution towards more integrated research on the Holocaust. This article will discuss our current efforts and present initial results in some key areas midway through the project. First we will present some of our innovations such as our identification work to discover relevant collections and archives and bringing their descriptions together into a single point of access: the EHRI portal. We will then dig deeper into the problem of the dispersion of Holocaust material by discussing a particular case study: the Terezín (Theresienstadt) ghetto's documentation, and the digital research guide that we are developing to address the specific challenges of the Terezín use case.

Background and State of the Art

EHRI did not have to start from scratch: organisations throughout the world have already done excellent work in collecting and saving documents, objects, photos, film and art related to the Holocaust. Therefore, on the most basic level, EHRI started by identifying and merging information on institutions which hold Holocaust-related archival material, the collection-holding institutions (CHIs). The EHRI identification work's starting points were the *Directory of Holocaust-related Archives* of the Conference on Jewish Material Claims Against Germany;³ the *Guide des archives sur la Shoah* of the Mémorial de la Shoah;⁴ the list of institutions with which Yad Vashem has worked and has copied material from.⁵ This compiled list was further completed with information from the United States Holocaust Memorial

2 See for instance Andrea Löw's recent book about the ghetto in Lodz (Löw 2006). More recently see Laura Jockusch's book about the early Jewish documentation projects (Jockusch 2012).

3 http://www.claimscon.org/archivist_forum/archive_search.asp

4 http://www.memorialdelashoah.org/b_content/getContentFromTopNavAction.do?navId=12

5 As a Consortium partner, Yad Vashem provided EHRI with internal overviews of institutions surveyed (Yad Vashem, internal documents).

Museum⁶ and aggregators such as national research portals on archival sources on the Second World War (e.g. <http://www.archievenwo2.nl/>, or <http://www.ns-quellen.at/>), national archival guides, experts and published studies on the subject.⁷ All the collected information on CHIs has been entered into a central repository. This repository uses the ICA-AtoM software, which is fully open source and complies with the ISDIAH standard.⁸

The screenshot displays the EHR I (European Holocaust Research Infrastructure) database interface. At the top, the logo 'E H R I' is shown next to the text 'EUROPEAN HOLOCAUST RESEARCH INFRASTRUCTURE'. Below this is a search bar labeled 'Advanced search' and a navigation menu with options: 'archival descriptions', 'authority records', 'archival institutions', 'functions', 'subjects', 'places', and 'digital objects'. A dark navigation bar contains 'Add', 'Manage', 'Import', and 'Admin' with dropdown arrows.

The main content area is titled 'View Archival Institution' and features a sidebar for 'Yad Vashem'. The 'Identity area' shows an EHR I identifier of 2798 and an authorized form of name 'Yad Vashem'. The 'Contact area' provides details for the primary contact, including an address in Jerusalem, Israel, and contact information for Haim Gertner. The 'Control area' lists the rules and conventions used (ISDIAH), the status (Draft), the level of detail (Partial), and the date of creation (Imported from EHR I spreadsheet at: 2012-03-09).

On the right side, there is an 'Upload limit' section indicating 0 GB of space for Yad Vashem, which is 'Unlimited'. Below that is a 'Holdings' section with a list of documents, including 'Bulletins of the Vaad Hahatzalah (Rescue Council) of the Jewish Agency in Eretz Israel, 1937-1959 (draft)', 'Documentation collected in preparation for the Eichmann Trial, 1933-1962 (draft)', and 'Documentation from Archives in Croatia regarding the Holocaust (draft)'.

Figure 1 Screenshot of the current EHR I database (work version)
(source: <http://icaatom.ehri-project.eu/index.php>)

6 mostly via the “Archival Guide to the Collections” (which provide a general overview of the collections of textual records available in the Museum’s Archives), and via the “Archival Finding Aids” (detailed inventories and finding aids that have been produced for selected collections in the Museum’s Archives) (www.ushmm.org)

7 An overview of the used archival guides will be published onto the EHR I portal in the course of 2013.

8 For general information ICA-AtoM we refer to <https://www.ica-atom.org/>. The EHR I version of ICA-AtoM has the following address: <http://icaatom.ehri-project.eu/>.

The next step of our general workflow is the identification of collections within the CHIs, which hold Holocaust-relevant material. All EHRI collection descriptions follow the ISAD (G) standard and will be – as much as possible – provided in English. The descriptions of these collections can be brought into the portal, and linked to the CHIs' information via two major pathways: harvesting or manual data entry. Using harvesting, the project would have the highest level of sustainability. However, in the first instance, identification will also happen manually, as many partner institutions currently do not provide a harvesting service. To have as many institutions as possible join the 'digital connection' group, a tool and example are being developed to help institutions to meet the future requirements of EHRI if they can and want to. Our surveying work, which we describe next, collects all this technical information next to its main focus on identifying sources.

In order to provide an overview that is as complete as possible, all sources related to the Holocaust could and should in principle find their place on the EHRI portal. EHRI does not want to exclude, but for pragmatic reasons some criteria of prioritization have to be put in place. For instance, EHRI has decided to prioritize the identification of victims' sources, because perpetrator sources are more numerous and less dispersed and therefore better known. Special attention is given to identifying Eastern (and Central) Europe sources, as these are the places where most victims perished, and where sources – compared to the West – have been less inventoried and made accessible.

EHRI's surveying work starts with Germany and its allies, and the countries occupied by the Axis Alliance, but EHRI will not stop here and further countries will get included in the further stages of the project, too. Any place where Holocaust-related sources can be found should in the end be included. In order to structure the identification work, EHRI developed a concise internal working definition of the Holocaust, which we see as a tool providing the EHRI consortium with a framework.

To complete EHRI's identification work within a systematic and structured framework, national reports are being written, which will give a short overview on a country's Holocaust history and its archives. All EHRI national reports follow the same general structure. First, in two short paragraphs a general overview of a country's history during the Second World War is given. The first paragraph covers questions of statehood as well as German rule and influence, while the second paragraph focuses on Holocaust history and also includes information on the size of the pre-war Jewish community as compared to the total population of a country and an estimated

number of Jewish victims. In a second section the reports describe briefly the archival situation. Again, the same structure is used for all national reports. A first paragraph deals with the archival culture of a country and how its archives are organized (centralized system, role of the state, and legislation can be addressed here). The second paragraph gives more information on which archives are most relevant for Holocaust research. The third part of the reports will give a concise overview of EHRI's research in the country. This results in a one to (at most) two pages executive summary per country. In consulting these summaries, the reader should get a clear and concise overview of why the country is being researched and what the current state of knowledge and access to Holocaust-related archives in this country is, including a concise overview of EHRI's identification efforts.

In doing all this, EHRI supports new research and enables historiographical progress by providing access and by sharing information on sources relatively unknown until now. EHRI's identification work assists the Holocaust studies' globetrotters so that they know as much as possible before they start planning travels. Researchers will be able to evaluate more easily which sources are located where, whether there are duplicates available in other repositories, etc.

Next to this more global overview of resources in national reports and collection descriptions, Holocaust studies is also supported by specific research guides giving detailed access to individual topics such as the Terezín ghetto.

Dispersion of Records and Challenges for Holocaust Research – The Terezín Case

The ghetto in Terezín (Theresienstadt in German) was one of the major places of suffering and death of Jews from Bohemia and Moravia, Germany, Austria, the Netherlands, Denmark and other European countries. Out of approx. 150,000 prisoners, over 30,000 died there between 1941 and 1945 due to starvation, overcrowded and unhygienic accommodation and diseases. Another 90,000 were deported to the ghettos and extermination camps in the East, from where only roughly 4,000 returned. Unlike most other ghettos in Nazi occupied East-Central Europe, Terezín was not liquidated at the end of the war. A fraction of its prisoners survived inside of the ghetto walls and were liberated in May 1945. The ghetto has been used for Nazi propaganda

purposes and served as destination for old Jewish people from Germany and Austria. In 1944–1945, in an attempt to mislead the world about the genocide of Jews, Terezín was showcased to delegations of the International Committee of the Red Cross.⁹

The Terezín ‘Council of Elders’, the Jewish ‘self-administration’ produced a large amount of documents, the lack of paper in the ghetto notwithstanding. However, most of these documents have been destroyed at the end of the war on the orders of the SS Commander of the ghetto. Karel Láguš and Josef Polák describe how all materials relating to the pre-1945 period were taken away and that the search for ‘dangerous’ documents extended not only to the offices of the ‘self-administration’, but also to the lodgings of the prisoners. Especially lists and card files of murdered people and of those who were deported to the ghettos and extermination camps in the East were confiscated (cf. Láguš/Polák 1964: 201), as was the physical evidence of death in Terezín: the ashes of people who perished in Terezín were taken away and partly dispersed in the Ohře/Eger river and partly put into an unmarked pit outside of Terezín.

The surviving original documentation was thus either actively used or created after 1st January 1945 (such as for instance a card file of prisoners who mostly survived in the ghetto until its liberation in May 1945), kept illegally by various groups of prisoners, or (re)created after the liberation. Some documents in the central registry or their copies such as the transport lists were hidden and survived its evacuation. A number of prisoners collected and saved various documents: for instance, Karel Herrmann (Heřman) who documented cultural life in the ghetto (cf. Kryl 1986; Štefaníková 2004) or H. G. Adler, one of the future historians of the ghetto. The most extensive set of documents was gathered by Hechalutz, the Zionist youth movement around Zeev Scheck, and was transported to Prague after liberation by his girlfriend Alisah (meanwhile, Scheck was deported to Auschwitz and eventually liberated in Dachau).

Terezín is known for the cultural production of prisoners – however, much of the art work and of prisoners’ diaries vanished with their authors following the deportation to extermination camps. Only some manuscripts or pictures were left with friends in Terezín or hidden in the walls or other hide-

⁹ Of the extensive, though not very recent, literature on Terezín, see mainly Lederer (1983), Adler (1955), Láguš/Polák (1964), Kárný (1991), and the periodicals *Terezienstädter Studien und Dokumente* as well as *Terezínské listy*.

outs. The Terezín resistance managed to save some important art works that testified about the reality of life in the ghetto, including those of a group of painters employed in the Technical Department who used their access to paper and other material to document the reality of Terezín and who were discovered and deported to the Small Fortress (a concentration camp-like Gestapo prison) in Terezín. Years after the liberation, documents were found in Terezín, for instance the diary of Egon Redlich (1995). On the other hand, in contrast to some of the other major Polish ghettos, very few authentic photos and film footage from the Terezín ghetto are available. The 1942 and 1944 propaganda films and the photos shoot around the filming in ‘beautified’ Terezín are to be used only with utmost caution (cf. Strusková 2011; Margry 1998). The authentic photo material consists mainly of a series of snapshots taken secretly by a Czech gendarme guarding the ghetto and an extensive series made in Terezín shortly after liberation.

Immediately after liberation, a group of Zionist activists led by Scheck started the *Dokumentační akce* (or ‘Documentation Project’), a Czech (and later Slovak) version of the Jewish documentation initiatives which collected documents and testimonies in many European countries.¹⁰ Within the short period between liberation and immigration to Palestine, the group collected testimonies, documents, photos and artwork documenting the persecution of Jews from Bohemia. In 1946, after Scheck had moved to Palestine, the collection was divided: the larger part was sent to Palestine and a smaller one was placed with the Jewish Museum in Prague. The division of material followed the arrangement submitted to the Jewish Agency, which sponsored the project: duplicate documents including most of the testimonies would go to the JMP and to Palestine, whereas originals would mainly be sent to Palestine. In a clear demonstration that documentation and immigration to Palestine, as well as the struggle for the Jewish state in Palestine, went hand-in-glove for Scheck and his colleagues, the boxes with documents also contained illegal content like weapons. At the same time, documents were collected in the Jewish Museum in Prague where H. G. Adler worked until 1947, when he finally emigrated to the United Kingdom. The collection of the *Dokumentační akce* and of the JMP was apparently used by the first historiographers of Terezín, Zdeněk Lederer (1983), Karel Lagus and Josef Polák (1964). H. G. Adler (1955) also drew on this material while working on his influential monograph.

10 For a comparative history of these projects, see Jockusch (2012).

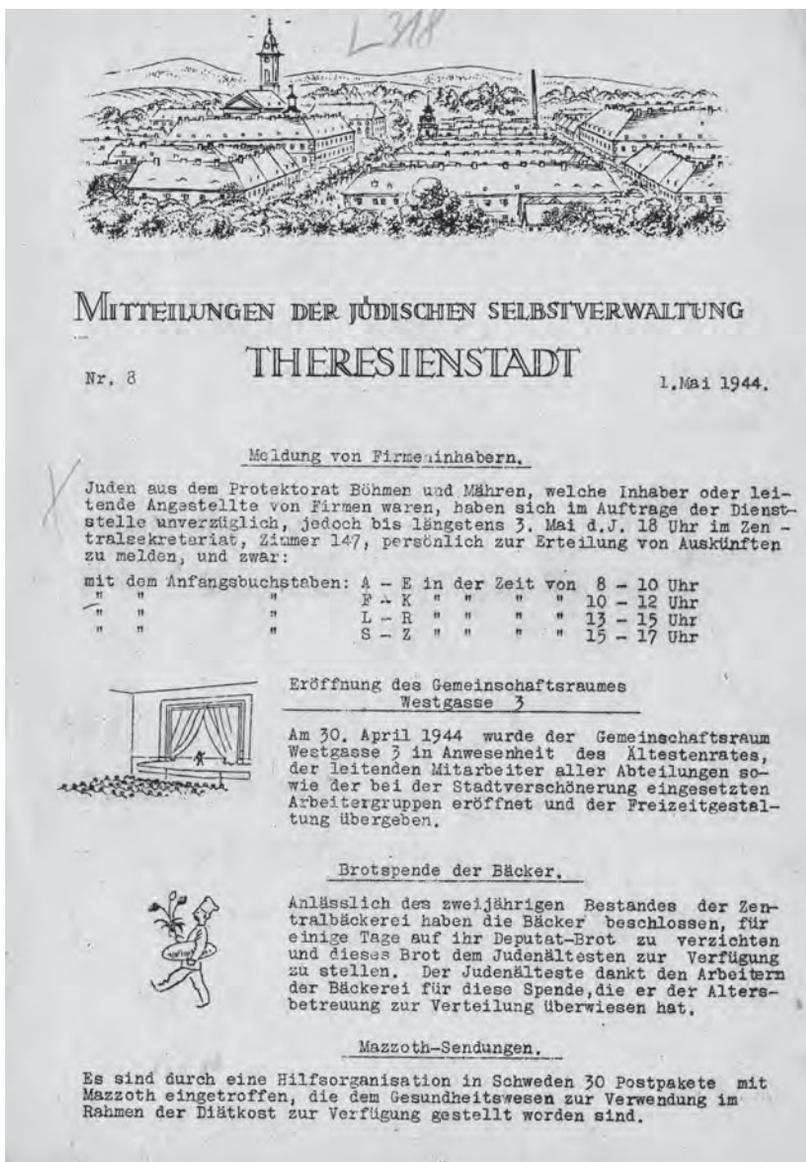


Figure 2 Daily bulletin of the Jewish self-administration, 1st of May 1944 (Archive of the Jewish Museum in Prague). Originally called orders of the day (Tagesbefehle), the daily bulletins were later renamed and graphically enhanced for the purposes of the Nazi propaganda. The bulletins, in originals or copies, are divided between all major Terezín archives.

As a result, the most important Terezín archives are located in dedicated memorial institutions. Following the Israeli War of Independence, the Documentation Project archive had been stored at the university campus at Mount Scopus, where it was not easily accessible. Only after the 1967 war was the collection moved to Yad Vashem and made available to researchers. Scheck also brought the Hechalutz Terezín collection to Palestine and donated it to the Central Archives for the History of the Jewish People, while Scheck continued to extend it. In 1976, most of the collection was transferred to Yad Vashem, whereas a smaller part was kept in the Central Archives and some materials, especially photos, were moved to Beit Theresienstadt in the Givat Haim kibbutz in Israel. Founded by Terezín survivors (including Zeev and Alisa Scheck) in 1960s, Beit Theresienstadt is a museum, an archive and an educational institution. The Terezín collection in Yad Vashem was later also extended by a collection of transport lists and albums devoted to the activity of various departments of the 'Council of Elders' which was saved by Hermann Weisz and acquired after his death in 1979. A third subcollection contains mostly personal information and documents provided by the former inmates and their families.

The Terezín collection in the Jewish Museum in Prague based on the Documentation Project as well as other materials and over time extended by further acquisitions, was organised later into a form that roughly corresponded to the structure of the Jewish 'self-administration' in Terezín. Therefore, the Terezín collection attempts to partially reconstruct the largely destroyed and fragmented original Terezín documentation. The collection continued to grow since the fall of Communism, especially in conjunction with other projects of the JMP. Some interesting documents were received as part of the oral history project of the museum, another within the public appeal to share documents and photos providing information about their deported and murdered neighbours.

The personal story and the documentation trail of the Dokumentační akce leads also to the history of Beit Terezín. The archive of Beit Theresienstadt houses especially documents donated by the members of the organisation, including important artwork and children newspapers. Very soon after the foundation of the Terezín Memorial in 1947 (originally as Memorial to the Suffering of the Nation), its archive and later 'Documentation Department' was created which collected documents from former inmates and found in Terezín, as well as testimonies of former prisoners. Therefore, any serious researcher of the ghetto has to conduct research at least in these four major

Terezín archives: Beit Terezín, Terezín Memorial, Yad Vashem and the Jewish Museum Prague. Further significant Terezín collections and documents can be found in other archives around the world, for instance in the National Archives in Prague (many of the Terezín related documents were digitised by the Terezín Initiative Institute and are partly accessible online at www.holocaust.cz), the Center for Jewish History in New York (especially in YIVO archives), the Wiener Library in London, NIOD in Amsterdam, the International Tracing Service in Bad Arolsen or the Institute of Contemporary History in Munich. EHRI aims to integrate these resources on Terezín in a dedicated research guide.

Intellectual Integration: Identification Work and Research Guides

The primary aim of the EHRI Terezín Research Guide is to create a comprehensive, innovative and easy to use guide on the dispersed and fragmented Terezín (Theresienstadt) archival material and to empower further research on the history of the ghetto. The Terezín research guide illustrates the primary *raison d'être* of EHRI – to connect collections spread in many archives and in more countries. Moreover, EHRI and specifically the research guides it develops demonstrate what a collaborative archival project can achieve and how archivists can redefine their tasks beyond providing physical access and creating finding aids restricted to the local collections. The guide does not aim to make the existing archives irrelevant by placing all information online, but to help researchers identify relevant sources and to connect and compare them to documents in other collections. The guide will function as a gateway to the Terezín archival resources and – as an increasing amount of digitised material appears online (in fact, all four major Terezín archives either have already digitised their collection or are in the process of doing so) – it will point to the respective public online catalogues. We also hope that the guide will catalyse further research in the history of Terezín. While much has already been written and published, there is still no recent synthetic monograph about its history, structure and function, which would integrate Terezín into the broader research on the Holocaust.

In more than one way, the research guides test the challenges that EHRI as a whole needs to meet. The cataloguing standards and data collected from the four major partners show significant differences. While Yad Vashem is a large archive with extensive staff, Beit Terezín is a very small organisation

with few staff and limited archival competence. The Terezín Memorial, as a museum funded by the Czech Ministry of Culture, follows museum standards and tends to catalogue documents as individual items. The Beit Terezín archive differentiates between a collection of originals and subject oriented files (which – in turn – often contain copies of the originals). While Yad Vashem only catalogues the material on the file level (with files often containing hundreds of pages), other partner archives provide much more detail information going down to individual documents. Whereas the Terezín finding aid of the Jewish Museum Prague is hierarchical and contains up to ten levels, Yad Vashem's uses subcollections and files, Beit Theresienstadt the file level only, and Terezín Memorial works with separate items. Moreover, the main four Terezín collections integrated in the guide contain a number of copies of items also found in the other.

The archives use different cataloguing systems and standards: whereas Yad Vashem deploys a commercial system with the possibility of standard-compliant export, the Jewish Museum runs an open source cataloguing system with a very flexible metadata schema. The Terezín Memorial stores its data in a simple custom database created without taking into account any set of standards. By the start of the project, Beit Terezín only had short textual descriptions of its files (stored in separate MS Word files). Exporting the information in a compatible format and integrating it into EHRI is no easy task and requires a great deal of mapping and data transformation.

The EHRI research guides team comprises archivists and other experts from the Jewish Museum in Prague, Yad Vashem, Terezín Memorial, and Beit Terezín and has designed a strategy to integrate this diverse material. The team has agreed to keep most of the original finding aids and other descriptive data intact and focus on the structured metadata such as keywords, places, etc. which will make it possible to connect items from different archives. At the very outset of the project, a limited set of principal keywords was defined and a detailed hierarchical list of departments of the 'Council of Elders' (the Jewish 'self-administration' in the ghetto) was created by the Jewish Museum Prague in cooperation with the other team members. Within the project, the partners improved their data and – as much as possible – used the pre-defined metadata.

The Jewish Museum Prague analysed the provided data and where needed created the mappings between the different sets of metadata of the partners. Much attention has been devoted to the geography of the Terezín ghetto. The Jewish Museum Prague created an authoritative list of locations (including

GPS coordinates) inside the ghetto (houses, barracks, crematorium, etc.), which makes it possible to link documents to specific places on the current or historical map of Terezín. In future, the data might be used by the visitors to the site of the former ghetto on their portable devices, as they walk through the town.

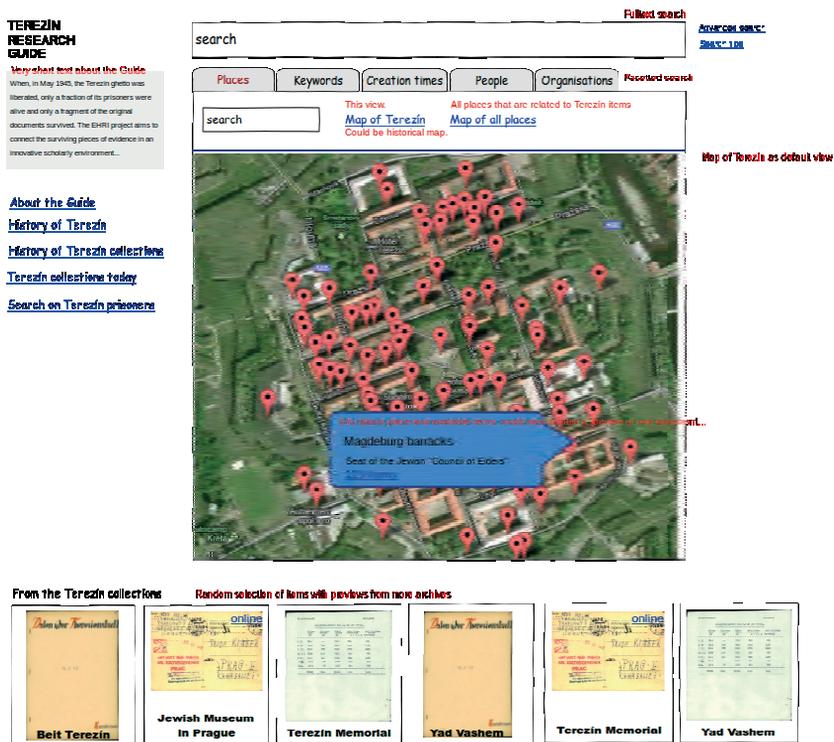


Figure 3 One of the early designs (wireframes) to approach the documents and files through the map of Terezín

Another principal type of metadata used to connect information and to make the work with the guide productive are databases of Terezín prisoners. The databases can provide a unified authoritative personal reference, as they contain further metadata for contextualisation and/or search. All four archives have access to the database created by the Terezín Initiative Institute in Prague (and in fact, the databases of the Jewish Museum Prague, Beit Terezín and Terezín Initiative Institute have been mapped onto each other),

but only the Jewish Museum directly employs this database as personal authority for cataloguing of the archive.

The guide will be designed not only for highly professional users (such as historians), but also with a view to the needs of students, interested members of the public or family members of Terezín prisoners. Therefore, the guides are meant to be used also by people who do not know how to work with traditional archival finding aids and have no extensive historical expertise about Terezín. Short biographic information was prepared for personalities most often referred to, as well as a basic timeline with details about different events and periods in the history of the ghetto. This contextual information will first help users formulate the query itself – for instance by offering information about people, definitions of keywords or descriptions of the functions of the main departments of the Terezín ‘self-administration’. On the other end of the search, the retrieved items (files or documents) will be contextualised not only by definitions of metadata, details about places or biographies, but also by placing the document on the chronological scale and making it easy to research related events and documents from the same period. A short history of Terezín, as well as the history of the principal Terezín archival collections, information about the databases of Terezín prisoners and other resources will be available as well.

The guide will differ from the more traditional archival guides or finding aids in that it will avoid ordering information along just one authoritative narrative or path, but will rather provide multiple ways and approaches. In the next year, we will be experimenting with ways which would enable researchers to traverse hierarchies of keywords, through the departments of the ‘Council of Elders’, and to approach the documents and files through the map of Terezín. The guide will be integrated into the EHRI portal and will share its collaborative features.

Technical Integration: The Central Point of Access

The aim of the EHRI online portal is to serve as a single point of access to descriptions of Holocaust-related archival sources that are scattered across hundreds of CHIs in and beyond Europe. It will bring together the results of the investigative work carried out elsewhere in the project (and discussed above), provide tools to support discovery and analysis of the integrated information as well as communication and collaboration among Holocaust

researchers and archivists. The portal will play a major role in turning currently disparate and fragmented Holocaust documentation into a cohesive corpus; thereby providing an online infrastructure that can enhance existing research, as well as inspire and facilitate new and innovative approaches to the history of the Holocaust.

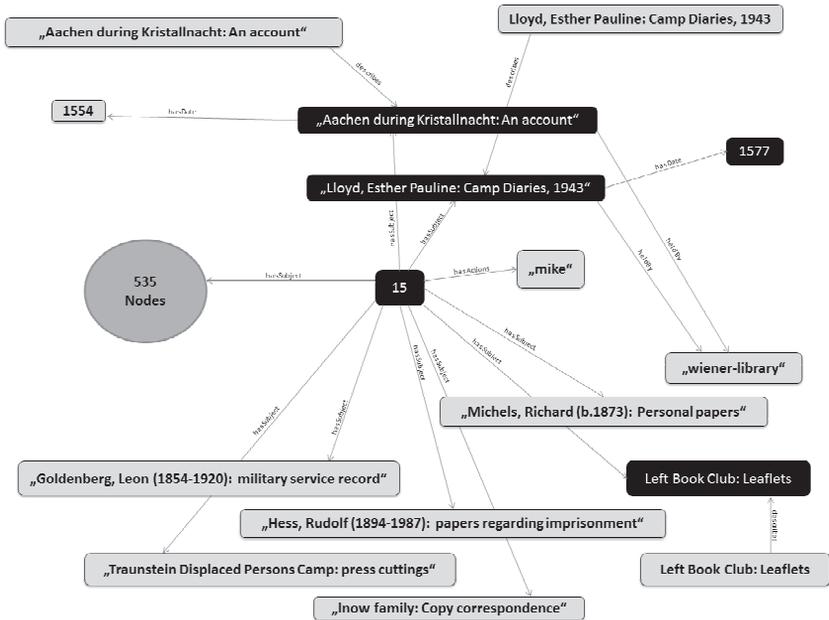


Figure 4 The EHRI archive graph

The backbone of the EHRI portal is an integration solution to the dispersed archival collections that are relevant to the study of the Holocaust. At the heart of this solution stands the EHRI metadata registry that will be filled with descriptions of relevant archival material harvested from CHIs. One of the main problems we face in this respect is that existing archival metadata is typically expressed in non-standardised and therefore highly heterogeneous forms. We address this problem in various ways such as harmonisation of metadata across sites, but we have also developed a new approach by using a graph database as the basis for our registry. Even though moving away from the tried-and-tested model of relational databases poses a (calculated) development risk, this is in our opinion more than counter-balanced by its benefits. Most importantly, a graph database is flexible with respect to information

schemas, and lends itself to the modeling of relationships between data elements in an adaptable and semantically rich fashion. Making best use of such capabilities, our metadata registry will facilitate inter-linkages between integrated descriptions, as well as the embedding of such descriptions in their wider archival and scholarly contexts.

Research users will interact with the information content of the portal via a variety of dedicated tools and services. Taken together, they will form a virtual observatory to enable discovery of, and access to Holocaust related archival sources. Just like astronomers require a (virtual) observatory, Holocaust researchers need a digital infrastructure to study the sources that are until now hidden and often locked away in archives. While we are currently still finalising the formulation of detailed user requirements for the virtual observatory, a few general remarks about some of its features are already possible:

Multi-lingual Search

We have shown above how Holocaust-related sources are dispersed across the globe, and how this dispersal requires researchers to negotiate a wide variety of languages, cultures and institutions. To complicate matters further, across countries and institutions, many different languages, perspectives and idioms have been brought to bear when cataloguing and describing these sources. This means that solely bridging across different metadata standards is not a sufficient integration strategy if truly unified access to the archival record of the Holocaust is to become a reality.

The EHRI portal will support unified access by incorporating a multi-lingual terminological database containing a dedicated thesaurus of well-defined subject terms relevant to Holocaust research and a set of authority files. The portal's search interface will use this database to translate and semantically expand user queries. This will enable users to locate all sources that are related to a given subject keyword regardless of differences in regard to provenance, language and/or employed descriptive idioms and terminologies.

Annotation Service

Description of archival material can by definition never be final, complete and exhaustive (cf., for instance, Duff/Harris 2002). Even though archives have invested heavily in producing descriptions that are as authoritative and rich as is possible, new research, or a new theoretical paradigm, may always

produce a new perspective on a given collection; bring to light connections between collections previously ignored; reveal factual mistakes in existing descriptions and so on. Archival descriptions should therefore not be fixed and monolithic, but 'living' and plural. EHRI will attempt to facilitate both by offering an annotation service.

In the context of the EHRI portal, research users will be enabled to annotate descriptions of archival material in a variety of ways. They may, for instance, use annotations to enhance a given description by attaching a textual note, or by linking it to a term from the EHRI thesaurus, an authority file, or another collection hosted by a different institution. Likewise, annotations could be used to embed an archival collection in its proper scholarly context by relating it to published research outputs such as journal articles, monographs, book chapters and so on. By experimenting with annotations, we are therefore attempting to bridge the gap between researchers and archivists for the benefit of both: researchers are empowered to actively contribute to archival descriptions, while archivists can make use of researchers' expertise and incorporate their suggestions in their own authoritative descriptions. The ultimate outcome should be rich and ever evolving descriptions illuminating the material from many different points-of-views.¹¹

Integrated Helpdesk

Another important EHRI service to strengthen ties between researchers and archivists is the integrated helpdesk. Such a service is crucially important. Indeed, our work on research user requirements suggests that a lot of Holocaust-related sources are currently found via direct contact between researchers and reference archivists rather than online searches.¹² By offering an integrated helpdesk, we plan to enhance this traditional and popular method of access in the context of the EHRI.

We are currently experimenting with a number of approaches taken from call routing (cf. Lee et al. 1998) and e-mail classifications systems (cf. Youn/McLeod 2007) to establish an automated helpdesk. To this effect, we are building a knowledge base containing semantic representations of archival

11 For a good discussion about using annotations and similar technologies to redefine the relationship between the archivist and the scholar and to improve access in an digital environment, see, for instance, Evans (2007), and Yakel/Shaw/Reynolds (2007).

12 That scholarly access to archival material is frequently mediated by a reference archivist has also been stressed by Pugh (1982), Duff/Fox (2006), and Johnson/Duff (2005).

institutions that hold Holocaust-related collections. In the final implementation, we are planning to give researchers the opportunity to submit their questions as free texts. The helpdesk system will extract relevant information from the texts, evaluate which of the institutions represented in the knowledge base is most suitable to answer, and reply with a list of institutions, ordered by an estimation of their suitability, and including further information that should help researchers to decide which institution to contact to discuss their problems in detail.

These and a host of further services will provide researchers with a set of useful tools to virtually explore the information content of the EHRI portal, to actively contribute to its development and to find their ways to relevant physical infrastructures. While we are therefore focusing on supporting scholarly research, we are also actively engaged in supporting the work of collection holding institutions. Unlike some of the large aggregators of Holocaust-related archival material which created physical copies of the files and largely separated them from the original archive, EHRI strives to connect researchers to the original institution and its expertise.

Apart from offering institutions the possibility to increase the visibility of their collections through integration of their metadata into the portal, we are currently investigating a number of services that could benefit archives. For instance, even though the portal will not hold digital copies of collections, EHRI may ultimately act as a broker to assist institutions with access to European wide services for data managements and storage. In a similar fashion, we are exploring the possibility of offering small institutions with Holocaust-related collections but no native capabilities to publish their findings aids online with an EHRI hosted archival description service. Finally, we are seeking to offer interested archives a range of Optical Character Recognition and Named Entity Recognition services to support and enhance present and future archival digitisation programmes (cf. Rodriguez et al. 2012).

The EHRI portal concentrates on integrating and enhancing existing capacities via a single point of online access, but by doing so also provides an environment to bring about change in both Holocaust research and archival theory and practice. The portal will offer Holocaust scholars the possibility to find relevant sources across institutional, national and linguistic boundaries and with opportunities to combine, contextualise, exchange and disseminate the results of their research in new ways. It offers CHIs tools and services that increase the online visibility of their collections, improve access to their

holdings, and enable innovative methods of engagement with their scholarly readers. The portal is, in other words, a central ingredient in an infrastructure that harnesses digital opportunities to connect people and resources in new ways.

Conclusion

This article started with a description of the current state of the art in Holocaust research and how it is characterized by many dispersed collections, partly distributed across continents. By connecting collections from different archives and countries with research, EHRI aims to address these important challenges that Holocaust researchers face and make a significant contribution towards more integrated research on the Holocaust. In the first two years of the project, EHRI has worked on the identification of institutions, which hold Holocaust-relevant collections. It has developed a dedicated workflow to survey these collections and has already delivered on national reports and overviews of archives in several European countries. But EHRI does not aim to just describe as many collections as possible but also to dig deeper into the analysis and description of especially challenging cases. A specific effort is dedicated in EHRI to linking sources related to the Terezín ghetto and developing an innovative research guide which will stimulate new research. All the results of the EHRI investigative work including the guide will be made available through a portal as a virtual central point of access. This portal is at the heart of the digital research infrastructure of EHRI.

As discussed, we identified (or we were able to identify) that setting up such a digital research infrastructure will have to overcome a number of problems that are directly connected to different archival cultures. However, we are confident that we have the solutions in place that will deliver a portal, which will connect people and resources in novel ways, with the aim of furthering our understanding of one of the most important events in European history. But a lot of research still needs to be done to find the best possible digital infrastructure for Holocaust sources, which are so dispersed and archived and published to different standards.

Literature

- Adler, H. G. (1955): *Theresienstadt 1941–1945. Das Antlitz einer Zwangsgemeinschaft: Geschichte, Soziologie, Psychologie*. Tübingen: J. C. B. Mohr (Paul Siebeck).
- Čapková, Kateřina, and Michal Frankl (2012): *Unsichere Zuflucht: Die Tschechoslowakei und ihre Flüchtlinge aus NS-Deutschland und Österreich 1933*. Köln: Böhlau.
- Duff, Wendy, and Allyson Fox (2006): You're a Guide Rather Than an Expert. Archival Reference from an Archivist's Point of View. In: *Journal of the Society of Archivists* 27, no. 2: 129–153.
- Duff, Wendy, and Verne Harris (2002): Stories and Names: Archival Description as Narrating Records and Constructing Meanings. In: *Archival Science* 2, no. 3: 263–285.
- Evans, Max (2007): Archives of the People, by the People, for the People. In: *American Archivist* 70, no. 2: 387–400.
- Jockusch, Laura (2012): *Collect and Record! Jewish Holocaust Documentation in Early Postwar Europe*. Oxford: Oxford University Press.
- Johnson, Catherine, and Wendy Duff (2005): Chatting Up the Archivist: Social Capital and the Archival Researcher. In: *American Archivist* 68, no. 1: 113–129.
- Judt, Tony (2005): *Postwar. A History of Europe Since 1945*. New York: Penguin Press.
- Kárný, Miroslav (1991): *Konečné řešení. Genocida českých židů v německé protektorátní politice*. Praha: Academia.
- Kryl, Miroslav (1986): A Significant Source of Information about Prisoners' Recitation and Theatrical Activities in the Terezín Concentration Camp-Ghetto". Karel Heřman's Collection Dating from the Years 1942–1945. In: *Judaica Bohemiae* XXII, no. 2 (1986): 74–86.
- Lagus, Karel, and Polák, Josef (1964): *Město za mřížemi* (Dokumenty; sv. 123). Praha: Naše vojsko.
- Lederer, Zdeněk (ed.) (1983): *Ghetto Theresienstadt*. New York: H. Fertig.
- Lee, C. H., Carpenter, R., Chou, W., Chu-Carroll, J., Reichl, W., Saad, A., and Zhou, Q. (1998): A study on natural language call routing. In: *Interactive Voice Technology for Telecommunications Applications. IVTTA '98*. Proceedings. 1998 IEEE 4th Workshop. Piscataway, NJ: IEEE Service Center, pp. 37–42.
- Löw, Andrea (2006): *Juden im Getto Litzmannstadt. Lebensbedingungen, Selbstwahrnehmung, Verhalten*. Göttingen: Wallstein.

- Margry, Karel (1998): Ein interessanter Vorgänger: Der erste Theresienstadt-Film (1942). In: *Theresienstädter Studien und Dokumente*, S. 181–212.
- Pugh, Mary (1982): The Illusion of Omniscience: Subject Access and the Reference Archivist. In: *American Archivist* 45, no. 1: 33–44.
- Redlich, Egon (1995): *Zítbra jedeme, synu, pojedeme transportem: deník Egona Redliča z Terezína 1.1.1942–22.10.1944* (Edice Knihy, dokumenty). Brno: Doplněk.
- Rodriguez, K. J., M. Bryant, T. Blanke, and M. Luszczynska (2012): Comparison of Named Entity Recognition Tools for raw OCR Text. In: *Proceedings of KONVENS 2012*, Vienna, 21 September 2012, pp. 410–414.
- Štefaníková, Jana (2004): Činnost Karla Herrmanna v Terezíně 1941–1944 (Heřmanova sbírka, a její osudy). In: *Terezínské studie a dokumenty* (2004): 62–117.
- Strusková, Eva (2011): Film Ghetto Theresienstadt: Die Suche nach Zusammenhängen. In: Ronny Loewy and Katharina Rauschenberger (Eds.): *Der Letzte der Ungerechten: Der Judenälteste. Benjamin Murelstein in Filmen 1942–1975*. Frankfurt: Campus, pp. 125–158.
- Yakel, Elizabeth, Seth Shaw, and Polly Reynolds (2007): Creating the Next Generation of Archival Finding Aids. In: *D-Lib Magazine* 13, no. 5/6. Online: <http://www.dlib.org/dlib/may07/yakel/05yakel.html>.
- Youn, S., and D. McLeod (2007): A Comparative Study for Email Classification. In: S. V. Nath (ed.): *Advances and Innovations in Systems, Computing Sciences and Software Engineering*. Dordrecht: Springer, pp. 387–391.

Erfahrungen aus dem Projekt „WissGrid – Grid für die Wissenschaft“

von Harry Enke und Bernadette Fritzsch

Kollaborationen über die Grenzen der eigenen Einrichtung hinweg gehören schon seit Langem zum Alltag eines Wissenschaftlers. Waren es in früheren Jahrhunderten vor allem Korrespondenzen zwischen führenden Gelehrten (vgl. z. B. Brockliss 2002) und später die Veröffentlichung in wissenschaftlichen Journalen, die neue Erkenntnisse in der Community zu verbreiten halfen, so eröffnen die jetzt vorhandenen Technologien neue Möglichkeiten der Zusammenarbeit. Daten, Methoden und Ergebnisse wissenschaftlicher Arbeit können in einem viel stärkeren Maße und vor allem zeitnah ausgetauscht werden, sodass Kollaborationen zwischen verschiedenen Forschungseinrichtungen, die über den ganzen Globus verteilt sein können, immer besser möglich werden.

Für die Nutzbarmachung der Grid-Technologie für den Aufbau kollaborativer Arbeitsumgebungen hat die vom BMBF geförderte D-Grid-Initiative¹ einen wichtigen Rahmen geschaffen. Im ersten Call hatten fünf akademische Communities die Vorreiterrolle übernommen und fachbezogene (und generische) Infrastrukturen für die gemeinsame Nutzung wissenschaftlicher Ressourcen entwickelt und prototypisch aufgebaut. Zu diesem Kreis gehörten die Hochenergiephysik, die Astronomie, die Klimaforschung, die Medizin und die Geisteswissenschaften. Diese schlossen sich zusammen, um im dritten Call der D-Grid-Initiative die bei der bisherigen Arbeit gewonnenen Erfahrungen in einem gemeinsamen Vorhaben für weitere akademische Communities nutzbar zu machen. Daraus entstand das Projekt „WissGrid – Grid für die Wissenschaft“². An ihm waren insgesamt 15 Partneereinrichtungen aus den verschiedenen Communities beteiligt. Durch diesen interdisziplinären Ansatz sollten Synergien zwischen den Gridprojekten ermittelt und genutzt werden, um neue Fachcommunities an die Gridtechnologie heranzuführen.

1 www.d-grid-gmbh.de

2 www.wissgrid.de

Insbesondere wurde WissGrid initiiert, weil sich in den ersten Jahren von D-Grid herauskristallisierte, dass die Anforderungen des akademischen Bereichs sich in vielen Aspekten von Grids in der Wirtschaft unterscheiden. Da die Projekte des zweiten und dritten Calls der D-Grid-Initiative vor allem auf die Einbindung der Industrie ausgerichtet waren, bildete WissGrid einen Gegenpol durch die Fokussierung auf den wissenschaftlichen Bereich und die Orientierung auf die Anforderungen der Grid-Nutzer.

Der Aufbau einer kollaborativen Arbeitsumgebung für die Wissenschaft umfasst nicht nur die Lösung technischer Probleme. Vielmehr ist auch eine ganze Reihe von nichttechnischen Fragestellungen zu bearbeiten, die den gesamten Arbeitsablauf des Wissenschaftlers betreffen. Die im Projekt gewonnenen Erfahrungen in technischen und nichttechnischen Bereichen sollen im Folgenden kurz zusammengefasst werden, wobei an geeigneter Stelle auf die im Projekt erstellten Dokumente (siehe Literaturangaben) verwiesen wird.

Vorarbeiten

Die Arbeiten von WissGrid bauen auf den vorangegangenen Community-Projekten auf, in denen spezifische, auf die jeweiligen Disziplinen zugeschnittene Lösungen erarbeitet wurden.

- *HEP-Grid (Hochenergiephysik)*

Die Grid-Aktivität der Hochenergiephysik ist durch um das CERN aufgebaute LHC Computing Grid dominiert. Im Unterschied zu den anderen Community-Grids ist hier der Entwicklungsprozess des Grids schon im Betriebsstadium. Die HEP-Community konnte sich daher auf Beiträge zu einem bestehenden, funktionierenden Grid konzentrieren, welches in seiner Existenz und Funktion kaum von den HEP-Aktivitäten des D-Grid beeinflusst wurde. Das LHC-Grid hat wenige großvolumige Datenquellen, deren Daten zu großen Teilen in einem gridbasierten Netz von Speicherknoten verfügbar sind, und für diese Strukturen existiert ein breit ausgearbeitetes System von Workflows mit entsprechenden Tools. Dies war einerseits ein großer Vorteil, hat jedoch anderen Teilen der Community z.T. nicht genug Entwicklungsspielraum für die Adaption der Grid-Technologie für ihre Zwecke geboten.

- *AstroGrid-D (Astronomie)*

Die Grid-Aktivitäten in der Astronomie sind parallel zum Aufbau des Virtual Observatory (VO) gewachsen. Während die Grid-Aktivität sich

auf den Aufbau von Storage- und Compute-Infrastruktur konzentrierte, hat das VO die Standardisierung von Datenformaten und Archiv-Strukturen zur Publikation von Daten vorangetrieben. Diese Arbeiten können als komplementär zueinander betrachtet werden. Auf Initiative des AstroGrid-D wurde eine wichtige Infrastruktur-Komponente, das Management der virtuellen Organisationen (basiert auf einem den Communities zugänglichen Webservice-System), im D-Grid eingeführt. Im Verlauf des Projekts hat sich der Schwerpunkt der Arbeiten von der Integration von Compute-Ressourcen zum Realisieren von verteilten Datenzentren verschoben. Die großen Datensammlungen der Astronomie (vgl. u. a. Enke/Wambsganz 2012) stammen aus vielfältigeren Quellen als in der Hochenergiephysik und folglich waren flexiblere Grid-Implementierungen notwendig. Aus diesem Grund hat sich ein zunächst angedachtes gemeinsames Grid mit der Hochenergie als nicht tragfähig erwiesen.

- *C3Grid (Klimaforschung)*

Die relevanten Daten der Klima-Community stammen aus sehr heterogenen Quellen: Neben Langzeitmessreihen an ausgewählten Standorten werden umfangreiche Messkampagnen bei Expeditionen durchgeführt. Zunehmend stehen satellitengestützte Messdaten zur Verfügung. Daneben bilden die Ergebnisse von Simulationen eine bedeutende Datenquelle. Diese unterschiedlichen Daten werden in zumeist spezialisierten Datenarchiven gespeichert, wobei jeweils spezifische Mechanismen zur Archivierung und Beschreibung eingesetzt werden. Das Collaborative Climate Community Data and Processing Grid (C3Grid)³ bietet für die unterschiedlichen Archive eine einheitliche Schnittstelle für die Suche und den Zugriff auf die Daten. Zusätzlich stellt das C3Grid verteilte Rechenressourcen zur Verfügung, um mit den Daten diagnostische Workflows auszuführen.

- *MediGRID (Medizin)*

Das MediGrid ist vor dem Hintergrund von Umstrukturierungsprozessen in der medizinischen Informatik-Infrastruktur mit dem Ziel gebildet worden, einige weitgehend schon ausgearbeitete Workflows in verschiedenen Bereichen (Genomforschung, Klinische Forschung, Bildverarbeitung) durch Einsatz von Grid-Computing effizienter zu gestalten. Als Einziges der Call-1-Projekte konnte sich MediGrid auch Förderung im

3 www.c3grid.de

zweiten D-Grid-Call⁴ sichern, da in diesem Bereich auch eine Kooperation mit kleinen und mittleren Unternehmen (KMU) leicht möglich ist. Die TMF e.V. als Dachorganisation der Medizinischen Verbundforschung hat für diese Grid-Projekte mit dem Forum Grid eine nationale Kooperationsplattform gebildet.⁵ Das Nutzen von verteilten Computere Ressourcen für dedizierte Workflows stellt für MediGrid die zentrale Komponente der Grid-Infrastruktur dar und wird über ein Portal zur Verfügung gestellt. Die gemeinsame Nutzung von Archiven und Datensammlungen haben in der Medizin wesentlich höhere Anforderungen an Datenschutz und -sicherheit und stand nicht im Fokus der Grid-Arbeiten.

- *TextGrid (Geisteswissenschaften)*

TextGrid startete mit dem Anliegen, „eine grid-fähige Workbench für die philologische Bearbeitung, Analyse, Annotation, Edition und Publikation von Textdaten“ zu erschaffen, und hat mit dem TextGridLab und dem TextGridRep zwei Komponenten einer „Virtuellen Forschungsumgebung für die digitalen Geisteswissenschaften“ implementiert. Das TextGridLab ist ein Webservice-basiertes Werkzeug, das einige der wichtigsten Workflows im geisteswissenschaftlichen Bereich organisiert und kollaboratives Arbeiten mit Texten ermöglicht. Das TextGridRep als Archiv für Texte und Arbeiten, insbesondere aus dem TextGridLab, hat eine verteilte Ressourcen-Infrastruktur und Grid-basierte Ressourcen konzeptionell eingebunden. Parallel zur Beteiligung an WissGrid ist TextGrid weiter aus dem BMBF-Fachreferat gefördert worden.⁶

Technische Aspekte

Die unterschiedlichen Workflows in den beteiligten wissenschaftlichen Communities spannen bereits eine recht beträchtliche Bandbreite von sehr unterschiedlichen Anforderungen an kollaborative Arbeitsumgebungen auf. Dementsprechend wurden von den Grid-Projekten auch unterschiedliche Ansätze für die jeweilige Grid-Infrastruktur verwirklicht. Für andere wissenschaftli-

4 services@medigrid, <http://services.medigrid.de/>. Neben dem MediGrid wurden im Rahmen der D-Grid-Calls außerdem MedinfoGRID (Call 2) und PneumoGRID (Call 3) gefördert.

5 http://www.tmf-ev.de/Ueber_uns.aspx

6 TextGrid ist aus dem Fachreferat des BMBF von Anfang an gefördert worden, hat sich jedoch im D-Grid-Beirat und bei D-Grid-Aktivitäten stark engagiert.

che Communities mit ihren spezifischen neuen Anforderungen ist in der Regel nicht zu erwarten, dass die bereits entwickelten Grid-Strukturen als Ganzes einfach übernommen werden können. Die in den jeweiligen Community-Grids gefundenen technischen Lösungen wurden daher in WissGrid zunächst einmal grundlegend analysiert. Unterschiedliche technische Lösungen für ähnliche Probleme wurden zusammengestellt und verglichen. Es entstand ein Pool von Einzelkomponenten, die für neue Communities dann je nach ihren Anforderungen neu zusammengestellt werden können.

In WissGrid wurden im Arbeitspaket 2 eine Reihe von Dokumenten erarbeitet, die unter dem Sammelbegriff „Blaupausen“ eine Zusammenstellung von Erfahrungen und technischen Lösungen zu verschiedenen Aspekten einer kollaborativen Arbeitsumgebung enthalten (vgl. Stöckle et al. 2011; Schlünzen 2011; Grimme 2011; Schultes 2011; Rapp/Schmidt 2011). Sie erlauben es, sich einen schnellen Überblick über die verfügbaren Technologien zu verschaffen und eine Vorauswahl für bestimmte Komponenten zu treffen. In einem weiteren Dokument wurden die unterschiedlichen Bedürfnisse der Communities nach Sicherheitskonzepten neu bewertet und analysiert (vgl. Grimme/Enke 2011). Die Nutzung der Grid-Technologie stand dabei im ersten Projektjahr im Vordergrund. Mit zunehmenden Erfahrungen in der Beratung erster Community-Vorhaben rückten jedoch die kollaborativen Umgebungen und deren Elemente zum Datenzugriff oder andere Workflows in den Mittelpunkt und die Grid-Technologie wurde verstärkt als ein Element unter anderen in der Arbeitsumgebung eingeordnet.⁷

Von größerer Bedeutung als die Blaupausen war jedoch die Etablierung eines Fachberaterteams, welches aus Spezialisten zu den verschiedenen Themenkomplexen aus den „alten“ Communities besteht und den neuen Communities für Anfragen und Beratung zur Verfügung steht. Hierdurch können sehr individuelle Lösungen erarbeitet werden, die viel präziser die Anforderungen erfüllen können. Dadurch steigt potenziell die Akzeptanz der aufzubauenden kollaborativen Forschungsumgebung in der neuen Community. In diesem Zusammenhang waren vor allem die direkten Interaktionen zwischen den Initiatoren einer kollaborativen Umgebung und den Fachberatern erfolgreich. Obwohl durch das Fachberaterteam eine Reihe von unterschiedlichen

⁷ Hierfür war auch die von der D-Grid GmbH organisierte „Expertise SOEB“ (vgl. Enke 2010) sehr wichtig. Die Studie stellt für WissGrid einen Meilenstein dar, insofern hier die Verschiebung des Ansatzes als notwendige Bedingung für eine sinnvolle Virtuelle Forschungsumgebung (VRE) erwiesen hat.

Möglichkeiten der Kommunikation angeboten wurden, die von Informationsworkshops über Mailinglisten bis hin zu Ticket-Systemen (z.B. das D-Grid-Ticket-System) reichen, wurden diese nur relativ wenig von neuen Communities genutzt. Diese Angebote werden erst dann wirksam, wenn bereits ein gemeinsamer Bezugsrahmen etabliert wurde. Die häufigsten Beratungssituationen waren jedoch im Vorfeld, bei der Herausbildung eines solchen Bezugsrahmens, notwendig. Als Konsequenz kann festgehalten werden, dass auch in Zeiten moderner Kommunikationstechniken die persönliche Präsenz als Ausgangspunkt für Beratungs-Aktivitäten bei der Heranführung neuer Communities an neue Technologien unersetzlich ist.

Bereits bei der Konzeption von WissGrid wurde ein ganzes Arbeitspaket⁸ den Techniken für die Bereitstellung von Forschungsdaten gewidmet, inklusive einer zumindest partiellen Implementation von Werkzeugen, die disziplinunabhängig eingesetzt werden können. Die Beratertätigkeit zeigte sehr deutlich den Bedarf, zusätzlich einen für die deutsche Forschungslandschaft angepassten „Leitfaden zum Forschungsdaten-Management“ (Ludwig/Enke 2011) zu erarbeiten. Dieses Dokument hat auch weit über das WissGrid-Projekt und dessen Beratungsarbeit hinaus Verbreitung und Anerkennung gefunden.

Forschungsumgebungen, die wichtige Workflows einer Disziplin organisieren sollen und eine entsprechende Komplexität aufweisen, existieren immer im Spannungsverhältnis zwischen den Wünschen der Wissenschaftler und der technischen Machbarkeit. In der Regel wird dabei auf Grundkomponenten Dritter aufgebaut. Im Grid ist dies die Middleware.⁹ Damit tritt bei einer grundlegenden Veränderung der Middleware-Technologie immer das Problem auf, dass diese Dienste neu angepasst werden müssen. Die in WissGrid vereinten Communities verwendeten als Basistechnologie in der Middleware fast ausschließlich Globus Toolkit¹⁰ (GT) 4.x, nur die Hochenergiephysik setzte auf gLite¹¹. Für den Übergang von GT4 auf GT5 waren

⁸ Anzumerken ist, dass sich im Laufe der Arbeiten auch eine Verschiebung des Schwerpunkts von den Techniken der Langzeitarchivierung hin zu der breiteren Fragestellung des Umgangs mit Forschungsdaten als sinnvoll erwiesen hat.

⁹ Im D-Grid wurden drei unterschiedliche Middlewares eingesetzt: gLite, Globus Toolkit und Unicore. Das D-Grid hat große Anstrengungen zu Integration dieser Middlewares unternommen.

¹⁰ www.globus.org

¹¹ glite.cern.ch

die notwendigen Anpassungen teilweise sehr aufwendig.¹² Es gibt nur wenige Dienste der Community-Grids, welche den Übergang zur neuen GT5-Version überlebt haben, vor allem diejenigen, welche nicht vollständig von der Grid-Middleware abhängig waren. Diese Beispiel verdeutlicht, dass auch nach dem Aufbau einer Infrastruktur für einen nachhaltigen Betrieb der Infrastruktur hinreichend personelle und finanzielle Kapazitäten bereitgestellt werden müssen, um jeweils notwendig Anpassungen vornehmen zu können und damit den Fortbestand einer Infrastruktur zu sichern.

Implikationen für den Wissenschaftler

Die Nutzung einer virtuellen Arbeitsumgebung greift tief in den wissenschaftlichen Prozess ein. Auch wenn in den vorgestellten Community-Grids zunächst versucht wurde, die bestehenden Arbeitsabläufe abzubilden und den Wissenschaftler bei umständlichen Schritten wie der Datenrecherche zu unterstützen, hat der Betrieb einer kollaborativen Arbeitsumgebung in jedem Fall Rückwirkungen auf die Arbeit der Wissenschaftler selbst. Die verwendeten Technologien haben derzeit noch nicht einen Stand erreicht, als dass man sie als „black box“ nutzen könnte. Auch wenn es zahlreiche Ansätze gibt, die Komplexität der Infrastruktur hinter einer Middleware zu verstecken und damit für den Nutzer transparent zu machen, zeigt doch die Erfahrung bei der Benutzung des Grids, dass zumindest ein gewisses Grundverständnis der zugrundeliegenden Technologie beim Nutzer notwendig ist. Dazu sollen beispielhaft einige wenige Aspekte beleuchtet werden.

Dieses Grundverständnis beginnt bereits bei den Limitierungen der Einzelkomponenten des Systems: Sind die Daten nicht auf Festplatten gespeichert, sondern in einem Bandarchiv archiviert, dauert der Zugriff auf die angeforderten Daten natürlich länger und das Anfordern von vielen kleinen Ausschnitten aus den Dateien ist ineffektiv. Darüber hinaus können einzelne Datenarchive unter Umständen mit einer geringeren Bandbreite an das Netzwerk angebunden sein und dadurch der Datentransfer eingeschränkt sein.

Da in der kollaborativen Umgebung in der Regel viele leistungsfähige Ressourcen miteinander verknüpft werden, ist eine funktionierende Sicherheitsinfrastruktur unabdingbar. Im D-Grid erfolgt dabei die Authentifizierung

12 Dieser Wechsel war v.a. deshalb sehr einschneidend, weil er mit einem Paradigmenwechsel der Implementation verbunden war; die Orientierung an Webservices wurde verworfen.

meist über Zertifikate, deren Handhabung vielen Nutzern bisher fremd ist. Sie müssen daher oft erst lernen, wie man Zertifikate beantragt, sie sicher aufbewahrt, sie verwendet und gegebenenfalls in unterschiedliche Formate konvertiert.

Für die Nutzung neuer Werkzeuge in der wissenschaftlichen Arbeit sind neben den bisher genannten technischen Aspekten weitere Kenntnisse wünschenswert. So sollte der Wissenschaftler ein Grundverständnis für die verwendeten Suchalgorithmen haben. Sie bestimmen, was als Ergebnis der Suche angezeigt (oder eben auch nicht angezeigt) wird und in welcher Reihenfolge Ergebnisse aufgelistet werden. Da diese Ergebnisse als Input für die wissenschaftliche Arbeit genutzt werden sollen, ist eine kritische Reflexion notwendig. Wie vollständig sind die gefundenen Ergebnisse? Wie vollständig können sie überhaupt sein, wenn die Infrastruktur flexibel ist und z.B. einzelne Server auch zeitweise nicht verfügbar sein können? Welche Auswirkungen hat dies auf die Reproduzierbarkeit von Ergebnissen? Ändern sich durch unterschiedliche Suchergebnisse die Schlussfolgerungen für die eigene wissenschaftliche Arbeit? Werden dadurch andere Thesen generiert? Sowohl die Nutzer der verteilten Arbeitsumgebung als auch die Bereitsteller müssen sich derartige Fragen immer wieder stellen. Die Nutzung der verteilten Arbeitsumgebung hat damit also Einfluss auf den Erkenntnisgewinn und die eigene wissenschaftliche Arbeitsweise.

Eine der Grundlagen moderner Wissenschaft ist das Veröffentlichen von Ergebnissen in wissenschaftlichen Publikationen und der dazu gehörige Peer-Review-Prozess durch die Community. Mit dem derzeit beobachtbaren Übergang zu datenzentrierter Wissenschaft (siehe *The Fourth Paradigm*; Hey/Tanskey/Tolle 2009) gewinnt die Publikation von Daten zunehmend eine ähnliche Bedeutung. Hierzu muss teilweise eine Kultur des „Datenteilens“ in den Communities weiterentwickelt werden. Hierzu gehört auch eine standardisierte Dokumentation der Daten. In den bisher etablierten Gedächtnisinstitutionen wie den Bibliotheken hat sich in der Vergangenheit eine ausgeklügelte Systematik¹³ entwickelt, um das dort abgelegte Wissen schnell zugänglich machen zu können. Hier kann der Bereich der Forschungsdaten oft noch lernen. Dennoch existieren in einigen Bereichen bereits erfolgversprechende Lösungen zur Beschreibung von Forschungsdaten. So existiert mit

13 Als ein Beispiel sei hier die *Allgemeine Systematik für öffentliche Bibliotheken* (Lehmann 2003) genannt, die eine standardisierte Klassifikation der Bestände in Bibliotheken erlaubt.

ISO 19115 / ISO 19139 ein international standardisiertes Metadatenprofil für geografische Informationen, das in der Klimaforschung eingesetzt wird. Für astronomische Daten, die im Allgemeinen wenig kommerzielles Interesse haben, ist der – wirtschaftlich bedingte – Druck zur Standardisierung geringer. Größere Datensammlungen in der Astronomie sind fast immer in internationalen Projekten entstanden. Daher gibt es allgemein genutzte und anerkannte Standardformate.¹⁴ In der Medizin, der Biologie als auch in den Sozialwissenschaften existieren eine Reihe konkurrierender Systeme für die Standardisierung von Forschungsdaten und Metadaten.¹⁵ In anderen Wissenschaftszweigen ist man noch relativ weit am Anfang des Prozesses der Standardisierung von Daten und Metadaten. Unabhängig von der Community ist jedoch festzustellen, dass die Kenntnisse und Fertigkeiten vieler Wissenschaftler beim Umgang mit Metadaten oft hinter den Anforderungen zurückbleiben. Hier zeigt sich u.a. ein starker Bedarf nach entsprechenden Lehrinhalten in der universitären Ausbildung der Wissenschaftler.

Kollaborative Umgebungen erlauben jedoch nicht nur, Daten über Instituts- und Fachwissenschaftsgrenzen hinweg mit anderen Wissenschaftlern zu teilen, sondern auch die dazu gehörigen Werkzeuge. Auch hierbei muss eine entsprechende Kultur in den Disziplinen oft erst noch entwickelt werden. Obwohl von ähnlicher Natur wie beim Datenaustausch sind die Probleme hier vielfach komplexer. Die Entwicklung von adäquaten Beschreibungen von Software und Werkzeugen steht noch ganz am Anfang¹⁶ und wird sicher in den kommenden Jahren an Bedeutung gewinnen. Die Einbeziehung von Software in das Datenmanagement und die Archivierung ist ein noch wenig erschlossenes Thema.

14 Der Standard für das Flexible Image Transport System (FITS) ist von der International Astronomical Union (IAU) festgeschrieben. Sein Ursprung liegt in den ersten Satellitenprojekten der NASA.

15 Diese Situation entsteht nicht zuletzt aus der Tatsache, dass die Messtechnik und Datennahme in vielen dieser Bereiche von kommerziellen Produkten bestimmt wird, deren Kalibrierung und Formate weitgehend proprietär bleiben – und damit auch die Möglichkeit der Einflussnahme auf Datenformate weit geringer ist. Es geht hier auch nicht um Metadaten-Standards, die aus dem Bereich der Bibliotheken entstanden sind und die wesentlich die generischen Metadaten-Komponenten von Publikationen adressieren.

16 Ein Beispiel liefert WSDL, eine Beschreibungssprache für Webservices: <http://www.w3.org/TR/wsdl>.

Die Nutzung von kollaborativen Arbeitsumgebungen kann die Wissenschaftler in ihrer täglichen Arbeit unterstützen und die Bearbeitung bestimmter Fragestellungen überhaupt erst ermöglichen. Gleichzeitig hat sie aber auch Einfluss auf die Arbeitsweise des Forschers und erfordert neue Kompetenzen und das kritische Hinterfragen bestehender Prozesse zum Erkenntnisgewinn.

Organisatorische und finanzielle Rahmenbedingungen

Der Aufbau und Betrieb einer kollaborativen Forschungsplattform stellt nicht nur in technischer Hinsicht eine Herausforderung dar. Vielmehr müssen dazu auch organisatorische und finanzielle Rahmenbedingungen geklärt werden.

Von grundlegender Bedeutung sind dabei der Kontext der Fachwissenschaft und das Vorhandensein von akzeptierten Organisationen innerhalb der Fachgemeinschaft. Eine Virtuelle Forschungsumgebung (VRE), die eine zu schaffende Infrastruktur errichten will, kann ihr Nutzungsprofil nur in Bezug auf eine (oder mehrere) Fachgemeinschaften definieren. Eine Zusammenarbeit mit bestehenden Fachorganisationen ist notwendig. Je weiter fortgeschritten der Prozess der Community-Bildung ist, desto höhere Nutzerzahlen sind für die VRE zu erwarten. Auch wenn die kollaborative Forschungsinfrastruktur zunächst vielleicht mit einer beschränkten Anzahl von Teilnehmern beginnt, so kann in einer gut organisierten Fachcommunity bereits in diesem Stadium weitgehend abgesichert werden, dass Anforderungen der ganzen Fachgemeinschaft zumindest in die strukturellen Architekturüberlegungen einfließen. Damit ist der Weg zu einer späteren Erweiterbarkeit leichter als in Fällen, wo die Community nur sehr lose organisiert ist und Architekturentscheidungen daher die Anforderungen von weiteren Gruppen nicht berücksichtigen.¹⁷ Dies führt dann vor allem beim Ausbau und nachhaltigen Betrieb zu Problemen.

Ein Beispiel einer bereits gut organisierten Fachwissenschaft ist die Hochenergiephysik. Die Kooperationen zwischen den einzelnen Einrichtungen sind im WLCG, dem Worldwide Large Hadron Collider (LHC) Computing Grid, eingebunden. Es umfasst europäische Institutionen rund um die Experimente am LHC sowie einen Zusammenschluss amerikanischer Wis-

¹⁷ Es kann dann auch die Situation eintreten, dass die VRE zu weit auseinanderliegende Anforderungen zu integrieren versucht und an deren Implementation scheitert.

senschaftsinstitutionen als „Open Science Grid“. Das WLCG basiert auf vereinsähnlichen Strukturen, die für jeweils eine Dauer von drei Jahren definiert werden. Grundlagen der Strukturen und Festlegungen der Rechte und Pflichten der Kooperationspartner werden in einem Memorandum of Understanding definiert. Die Teilnehmer eint das Ziel, die für die Auswertung von Experimenten benötigten großen Rechen- und Speicherkapazitäten der Kooperationspartner als Grid-basierte Infrastruktur übergreifend nutzen zu können und die Entwicklung von Auswertungssoftware voranzutreiben.

Häufig kann die Entwicklung und der Aufbau einer VRE durch projektbezogene Förderung finanziert werden. Beispiele hierfür sind die Community-Grids, die im ersten Call der D-Grid-Initiative aufgebaut wurden. Jedoch reicht eine solche Anschubfinanzierung allein nicht aus. Der dauerhafte Betrieb einer Infrastruktur muss garantiert werden. Die dazu notwendigen Kosten umfassen sowohl Mittel für den Betrieb der Ressourcen (Bereitstellung der Hardware, Stromkosten, Lizenzen, ...) als auch Personalkosten für die Betreuung der Komponenten. Weiterhin sind durch die Abhängigkeiten von externen Softwareprodukten immer wieder auch grundlegende Arbeiten an der Infrastruktur notwendig. Gleichzeitig entwickeln sich die Anforderungen der Nutzer weiter, sodass neue Dienste nachgefragt werden. Dafür muss eine solide Finanzierung gefunden werden.

Für das D-Grid bestand der organisatorische Rahmen konzeptionell aus Community-Grids und dem Integrationsprojekt (DGI), welches hauptsächlich aus den HPC-Zentren und einigen anderen Rechenzentren und Providern (z.B. DFN) gebildet wurde. Aufgabe des DGI, welches über die gesamte Zeit gefördert wurde, war die Bereitstellung von zentralen Infrastruktur-Komponenten und die Sicherung von Community-übergreifenden Aufgaben. Als Ergebnis sollte die Etablierung von Grid-Middleware als Zugang zu Ressourcen, die aus öffentlichen Mitteln gefördert werden, langfristig mit gesichert werden. Während die zentrale Infrastruktur bereitgestellt werden konnte, ist die Etablierung von Grid-Middleware an den Rechenzentren nicht einmal für die Ressourcen-Provider im akademischen Umfeld erfolgreich gewesen. Mit dem Ende des D-Grid bleibt derzeit nur eine im Rahmen der Gauss-Allianz angedachte Option, über Unicore als Grid-Middleware¹⁸ einen

18 Diese Middleware war und ist für die meisten Community-Grids nicht Basis von eigenen Komponenten.

Teil von Rechenzeit als Grid-Ressourcen anzubieten.¹⁹ Von einer Etablierung eines Grids zur effizienteren flächendeckenden Nutzung der vorhandenen Ressourcen ist hier nicht zu sprechen. Vor diesem Hintergrund werden Nachhaltigkeitsüberlegungen in Virtuellen Forschungsumgebungen mit Grid-Komponenten erheblich erschwert.

In WissGrid wurde die existierende Förderlandschaft analysiert, um zu evaluieren, wie ein Betriebsmodell für eine VRE aussehen kann (vgl. u. a. Dickmann/Sax 2011). Diese Analysen zeigen, dass die Nachhaltigkeit einer VRE derzeit nur dann gesichert werden kann, wenn es möglich ist, die VRE in das Profil von bestehenden Institutionen einzubinden, die die Grundfinanzierung übernehmen. Da sich jedoch oft die VREs als übergreifende Organisationen herausbilden, gerade weil die institutionellen Wege unzureichend sind, ist diese Lösung unbefriedigend oder nicht gangbar. Auch die Gründung einer neuen Organisation (Verein oder Unternehmung) ist nur selten machbar.²⁰ WissGrid hat für den akademischen Bereich einige Überlegungen mit den Förderern erörtert.²¹ Derzeit gibt es eine Vielzahl von Stellungnahmen und Lösungsvorschlägen, die jedoch meist die Notwendigkeit von Änderungen in den IT-Infrastrukturen für die Wissenschaft auf einem sehr hohen Level adressieren. Ein Beispiel hierfür sind die Empfehlungen des Wissenschaftsrats zu Informationsinfrastrukturen. Die gegebenen Anstöße bleiben jedoch oft im Dickicht der verschiedenen Interessen der beteiligten Institutionen und Verbände stecken. Eine aus den Arbeiten von WissGrid deutlich hervorzuhebende Erkenntnis ist, dass die Beteiligung der Communities an dem Dialog über die IT-Infrastruktur dringend verbessert werden sollte und VREs hierzu einen effizienten Weg weisen. WissGrid hat in seinen Untersuchungen auch aufgezeigt, wie durch Nutzung der Grid-Middleware eine Effizienz-Steigerung des Ressourceneinsatzes erfolgen kann, wenn auf Basis

19 Diese Rechenzeit ist ohnehin für die Nutzung durch Wissenschaftler aus diversen Communities vorgesehen, es wird nur auf einen anderen Verteilungsmodus übergegangen.

20 Wissgrid hat sich von vornherein auf die Interessen der akademischen Beteiligten im D-Grid konzentriert.

21 Siehe hierzu <http://www.wissgrid.de/veranstaltungen/abgeschlosseneveranstaltungen/expertenrunde1.html> und http://www.wissgrid.de/publikationen/presentations/Dickmann_2011-07-13.pdf sowie die Materialien des Abschlussworkshop von WissGrid (s. <http://www.d-grid-gmbh.de/index.php?id=171&L=0>) bei der D-Grid-Ergebniskonferenz 2012.

einer verteilten und vernetzten Infrastruktur die Rechenzentren durch Spezialisierung bei Beibehaltung der notwendigen Diversität die Anforderungen der akademischen Communities erfüllen sollen.

Zusammenfassung

- VREs bauen zwar (teilweise) auf gleicher technologischer Basis auf, die konkrete Umsetzung ist aber sehr fachspezifisch und individuell. Dabei lässt die inhaltliche Nähe des Forschungsgegenstands zwischen zwei Communities noch keine Rückschlüsse auf die notwendigen Komponenten zu. Hier sei auf die SOFI-Expertise verwiesen, die festhält, dass Anforderungen für diese VREs an die Datenhaltung mit der zertifikatsbasierten Grid-Infrastruktur besser erfüllt werden als mit einem rollenbasierten System, wie es beispielsweise von TextGrid genutzt wird.
- VREs benötigen einen hohen Beratungsaufwand. WissGrid hat mit den Blaupausen Dokumente zur ersten Orientierung für neue akademische Communities erstellt. Wichtiger und wesentlich effizienter ist aber persönliche Beratung durch Fachberater, die über Erfahrungen mit den vorhandenen Angeboten verfügen oder den Aufbau von VREs bereits mitgestaltet haben. Ein breiter Erfahrungshorizont des Teams aus verschiedenen Fachrichtungen ist sehr vorteilhaft.
- Die Nachhaltigkeit von VREs ist nicht allein aus den „Bordmitteln“ der Projekte oder der beteiligten Institutionen zu sichern. Die bestehende Förderstruktur bedarf einer teilweisen Modifikation, um langfristig Gelder für Support und die Anpassung der Infrastruktur an gewandelte Anforderungen der akademischen Forschung bereitzustellen. Die Herausbildung von VRE ist eine Reaktion auf Lücken und Mängel in der bestehenden IT-Infrastruktur. Über eine VREs ist es den Fachgemeinschaften möglich, sowohl neue Anforderungen an die IT-Infrastruktur zu formulieren als auch deren Implementierung zu realisieren. Die VREs können auch den Nutzern von bestehenden Infrastrukturen mehr Gewicht bei Entscheidungsprozessen verschaffen.
- Im Zusammenhang mit VREs kommt der Organisation des Zugangs zu Forschungsdaten zentrale Bedeutung zu. Forschungsdaten sind die Basis des Erkenntnisgewinns (soweit es sich um IT-gestützte Forschung handelt). Während es für wissenschaftliche Publikationen wohlbekannte Wege und Verfahren zur Archivierung gibt, die jedoch auch einem „digitalen Wandel“ unterliegen, hat für die Archivierung und Publikation von

Forschungsdaten diese Entwicklung erst begonnen. Aus den Arbeiten von WissGrid ist jedoch klar, dass für Forschungsdaten eine direkt von der Fach-Community kontrollierte und verwaltete Form der Datenpublikation und Archivierung die größte Effizienz aufweist. WissGrid hat mit dem Leitfaden zum Forschungsdaten-Management und der zugehörigen Checkliste für den deutschsprachigen Raum ein wichtiges Instrument für die Weiterentwicklung des Forschungsdatenmanagements geschaffen.

Literaturverzeichnis

- Brockliss, L. W. B. (2002): *Calvet's Web: Enlightenment and the Republic of Letters in Eighteenth-Century France*. Oxford: Oxford University Press.
- Dickmann, F.; Sax, U. (Hrsg.) (2011): Aspekte der Implementierung eines Betriebsmodells für virtuelle Forschungsumgebungen, WissGrid Deliverable 1.5. http://www.wissgrid.de/publikationen/deliverables/wp1/WissGrid_API_D15_v1_01_final.pdf.
- Enke, H.; Harms, P.; Dickmann, F. (2010): Expertise zur Errichtung einer Virtuellen Arbeitsumgebung fuer die Sozial-Oekonomische Berichterstattung (SOEB). http://www.wissgrid.de/publikationen/Expertise_VirtAug.pdf.
- Enke, H.; Wambsganß, J. K. (2012), „Astronomie und Astrophysik“, in: Neuroth, H. et al. (Hrsg.): *Langzeitarchivierung von Forschungsdaten: Eine Bestandsaufnahme*. Boizenburg: Hülsbusch, S. 275–293. Online: http://nestor.sub.uni-goettingen.de/bestandsaufnahme/kapitel/nestor_bestandsaufnahme_014.pdf.
- Grimme, C. (2011): Leitfaden zur Unterstützung des Software-Management-Prozesses, WissGrid Deliverable 2.1.6. <http://www.wissgrid.de/publikationen/deliverables/wp2/Software-Lifecycle-Management-1.01.pdf>.
- Grimme, C.; Enke, H. (2011): Sicherheitsinfrastrukturen in D-Grid, WissGrid Deliverable 2.3.2. <http://www.wissgrid.de/publikationen/deliverables/wp2/Sicherheits-Infrastrukturen-in-DGrid.pdf>.
- Hey, T.; Tansley, S.; Tolle, K. (2009): *The Fourth Paradigm*. Redmond: Microsoft Research.
- Lehmann, Kathrin (Bearb.) (2003): *Allgemeine Systematik für öffentliche Bibliotheken (ASB), Gliederung und Alphabetisches Schlagwortregister*. Erarbeitet vom Ausschuss für Systematik beim Verband der Bibliotheken des Landes Nordrhein-Westfalen. Bad Honnef: Bock & Herchen.
- Ludwig, J., Enke, H. (Hrsg.) (2011): Leitfaden zum Forschungsdaten-Management, WissGrid Deliverable 2.3.1. <http://www.wissgrid.de/publikationen/deliverables/>

wp3/WissGrid-oeffentlicher-Entwurf-Leitfaden-Forschungsdaten-Management.pdf (revidierte Druckausgabe [2013]: Glückstadt: Hülsbusch).

Rapp, A.; Schmidt, O. (2011): Blaupause für die organisatorische Struktur eines neuen Community Grid, WissGrid Deliverable 2.1.3. <http://www.wissgrid.de/publikationen/deliverables/wp2/CG-Organisationsstrukturen-V2.0.pdf>.

Schlünzen, F.; Agapov, I. (2011): Evaluation und Dokumentation existierender Architekturkonzepte, WissGrid Deliverable 2.1.5. <http://www.wissgrid.de/publikationen/deliverables/wp2/Gridarchitekturen-im-Ueberblick-V2.0.pdf>.

Schultes, J.; Sandvoss, S. (2011): Kosten-Nutzen-Abschätzung des Aufbaus von neuen Communities, WissGrid Deliverable 2.1.4. <http://www.wissgrid.de/publikationen/deliverables/wp2/Kosten-Nutzen-CG-Aufbau-V2.0.pdf>.

Stöckle, G.; Enke, H.; Grimme, C.; Schintke, F.; Schlünzen, F.; Lohmeier, F.; Pempe, W. (2011): Evaluation, Dokumentation und Registrierung von höheren Diensten, WissGrid Deliverable 2.1.2. http://www.wissgrid.de/publikationen/deliverables/wp2/Community-Dienste_in_D-Grid_2011-03-14.pdf.

Lehre und Qualifizierung

Qualifizierung im Bereich digitaler Langzeitarchivierung

von Achim Oßwald und Stefan Strathmann

Qualifizierung als Thema der Aktivitäten der Abteilung F&E

Qualifizierung ist ein Bestandteil der Aktivitäten der Abteilung F&E, der diese schon lange begleitet, mehr und mehr an Bedeutung gewonnen hat und in den letzten Jahren immer weiter ausgebaut wurde.

Ausgehend von den Aktivitäten im nestor-Projekt und den dort gemachten Erfahrungen hinsichtlich der Notwendigkeit und den Möglichkeiten der Qualifizierung, wurde dieses Thema in verschiedene andere Projektkontexte und darüber hinausgehende Aktivitäten integriert. Einerseits wurden von der Abteilung F&E verschiedene Fortbildungsangebote entwickelt und implementiert, die sich an interne Adressaten in der Abteilung und der SUB richten. Zum anderen wurden die eher nach außen gerichteten qualifizierungsbezogenen Projektaktivitäten deutlich verstärkt.

Zu den internen Schulungsaktivitäten zählt insbesondere die Reihe der F&E-Kolloquien, in der KollegInnen Zwischenstände und Ergebnisse ihrer Projektaktivitäten vorstellen, um darüber im Kollegenkreis konstruktiv zu diskutieren. Die Reihe ist sehr erfolgreich und die Veranstaltungen werden seit mehreren Jahren in etwa monatlichem Abstand durchgeführt. Hinzu kommen Einzelveranstaltungen zu unterschiedlichen Fragestellungen (Langzeitarchivierung, Social Media etc.), die verschiedentlich für die Mitarbeiter der SUB organisiert wurden.

Ebenfalls zu den internen Fortbildungsmaßnahmen kann die regelmäßige Betreuung von PraktikantInnen und ReferendarInnen gezählt werden, die in den letzten Jahren verstärkt in der Abteilung F&E für längere Zeiträume hospitierten. Gleichzeitig werden die Mitarbeiter der Abteilung ermuntert, sich regelmäßig fortzubilden. Hervorzuheben ist insbesondere, dass vermehrt von den verschiedenen Möglichkeiten bibliotheks- und informationswissen-

schaftlicher Fernstudiengänge Gebrauch gemacht wird.¹ Die so erlangten Erfahrungen und Kenntnisse fließen dann direkt wieder in die Tätigkeiten in der Abteilung ein.

Neben diesen eher nach innen gerichteten Qualifizierungsaktivitäten wurde das Thema Qualifizierung in den letzten Jahren immer mehr zu einem regelmäßigen Teil der Projektaktivitäten der Abteilung F&E.

So wird beispielsweise im Projekt DigCurV² gemeinsam mit verschiedenen europäischen und amerikanischen Partnern ein Rahmencurriculum für die berufliche Weiterbildung im Bereich der digitalen Langzeitarchivierung entwickelt. Das Projekt wird durch das „Leonardo da Vinci“-Programm³ der Europäischen Kommission zur beruflichen Bildung gefördert. Neben diesem Projekt, das sich speziell weiterbildungsspezifischen Fragestellungen widmete, engagiert sich die Abteilung in den jeweiligen qualifikationsspezifischen Arbeitspaketen verstärkt auch in einer Reihe von anderen Projekten. Die Einrichtung dieser Arbeitspakete und das Engagement in diesen unterstützt maßgeblich den Transfer der jeweiligen Projektergebnisse in die interessierten Communities. So hat sich die Abteilung F&E in der Vergangenheit bspw. in den entsprechenden Arbeitspaketen der Projekte nestor, DPE⁴ und SHAMAN⁵ engagiert und ist gegenwärtig in entsprechenden Arbeitspaketen der Projekte DASISH⁶, TextGrid⁷, DARIAH-DE⁸ und PERICLES⁹ eingebunden.

1 Vgl. bspw. den MALIS-Studiengang der FH Köln (s. <http://malis.fh-koeln.de/>) oder den Masterstudiengang Bibliotheks- und Informationswissenschaft der HU Berlin (s. <http://www.ibi.hu-berlin.de/studium/fernstudium>).

2 „DigCurV“ ist das Akronym für „Digital Curator Vocational Education Europe“, vgl. <http://www.digcur-education.org/>.

3 Vgl. http://ec.europa.eu/education/lifelong-learning-programme/ldv_de.htm.

4 Vgl. „Digital Preservation Europe“, <http://www.digitalpreservationeurope.eu/>.

5 Vgl. „Sustaining Heritage Access through Multivalent Archiving“, <http://shaman-ip.eu/start>.

6 Vgl. „Data Service infrastructure for the Social Sciences and Humanities“, <http://dasish.eu>.

7 Vgl. <http://www.textgrid.de/>.

8 Vgl. „Digital Research Infrastructure for the Arts and Humanities – DE“, <http://de.dariah.eu/>.

9 Vgl. „Promoting and Enhancing Reuse of Information throughout the Content Lifecycle taking account of Evolving Semantics“, <http://www.pericles-project.eu/>.

Organisatorische Rahmenbedingungen der Qualifizierungskooperationen

Die erste Förderphase des Projektes nestor (2003–2006) mit dem Ziel, ein Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit digitaler Ressourcen in Deutschland¹⁰ aufzubauen, hatte das Thema Qualifizierung noch nicht explizit im Fokus. Dennoch erhielt die Qualifizierung zu den Fragen eines drohenden Verlusts digital vorliegender Informationen sowie über Probleme und Lösungsansätze der Langzeitarchivierung digitaler Daten im Projektverlauf hohe Priorität: Der im Rahmen des nestor-Projektes über das Informationsportal www.langzeitarchivierung.de dokumentierte Sach- und Forschungsstand zum Thema Langzeitarchivierung war und ist erstmalig in deutscher Sprache in der gebotenen Breite und für jedermann zugänglich.¹¹ Zugleich förderte dieses Informationsportal die Vernetzung durch Informationen über Projekte und Fachleute in diesem noch relativ neuen Spezialgebiet. Damit wurde die Website zu einer Fundgrube für Praktiker sowie für die interessierte Öffentlichkeit, aber auch für Lehrende.

Die nestor-Projektbeteiligten und hier federführend die Abteilung F&E der SUB haben in dieser ersten Phase mit einer Reihe von Seminarveranstaltungen die Grundproblematik sowie den aktuellen Stand der Problemlösungsangebote zielgruppenspezifisch thematisiert und dokumentiert – weit aus umfangreicher und vielfältiger, als dies sonst bei Forschungsprojekten der Fall ist. Von all dem profitierten die weiteren Aktivitäten zur Aus-, Fort- und Weiterbildung im Bereich Langzeitarchivierung (LZA) immens: Angeregt durch entsprechende Bestandsaufnahmen und Forderungen aus dem Hochschulbereich (vgl. Oßwald/Scheffel 2006; Oßwald/Scheffel 2007), wurde am Ende der Förderphase I von nestor konsequenterweise im sog. „Me-

10 „nestor“ steht für „Network of Expertise in long-term **ST**Orage and availability of digital **R**esources in Germany“, vgl. <http://www.langzeitarchivierung.de>.

11 Vgl. für das gesamten Unterkapitel Scheffel/Oßwald/Neuroth (2010), darin insbesondere Kapitel 19.1 sowie die entsprechenden Vorgängerversionen dieses Handbuch-Abschnitts unter <http://nestor.sub.uni-goettingen.de/handbuch/artikel.php?id=75>, in denen die genannten Sachverhalte ausführlicher dargestellt werden. Die syntaktisch modifizierte Übernahme von Textstellen erfolgte in Abstimmung und mit Zustimmung der Koautorinnen dieser Beiträge. Der Verzicht auf wörtliche Zitate erfolgte mit Blick auf die Lesbarkeit des Gesamtbeitrags.

morandum zur Langzeitverfügbarkeit digitaler Informationen in Deutschland“ auch eine Aussage zum Thema Qualifizierung getroffen:

„18. Mit der digitalen Langzeitarchivierung entstehen neue Aufgaben für die archivierenden Institutionen. Es muss professionelles Personal zum Einsatz kommen. Die Anforderungen und Aufgaben der digitalen Langzeitarchivierung sind als ein Schwerpunkt in die Aus- und Fortbildung einzubeziehen. Gezielte Fortbildungsangebote sollten sowohl themenspezifisch sensibilisierend wie auch konkret qualifizierend angelegt werden.“ (nestor-Memorandum 2006)

Ziel der zweiten Phase des nestor-Projektes 2006–2009 war es daher, hierfür mittelfristig neue Angebote zu konzipieren.¹² Dazu wurde ein Arbeitspaket „Einrichtung und Ausbau von Ausbildungs- und Fortbildungsangeboten“ (AP 5) realisiert, das unter der Koordination und Leitung der SUB Göttingen vielseitige Aktivitäten entfaltete. Als besonders produktiv stellte sich die im Rahmen des AP 5 initiierte Arbeitsgruppe (AG) „Kooperation mit Hochschulen im Bereich Aus-, Fort- und Weiterbildung“ heraus, die sich ab 2007 zum Kristallisationskern der LZA-bezogenen Qualifizierungsaktivitäten entwickelte.

Bei den Qualifizierungsanstrengungen sind zwei Aktivitätsbereiche erkennbar:

- projektbasierte, von der nestor AG initiierte und realisierte Qualifizierungsangebote sowie
- hochschulbasierte Qualifizierungsangebote im Rahmen einschlägiger Curricula.

Ziel der Aktivitäten der von der Abteilung F&E der SUB Göttingen koordinierten nestor-AG „Kooperation mit Hochschulen im Bereich Aus-, Fort- und Weiterbildung“ war es, beide Aktivitätsbereiche strukturell zu stärken und zu verbinden, sodass im Sinne der Nachhaltigkeit die projektbasierten Aktivitäten weitergeführt, auf jeden Fall aber das im Laufe der Projektzeit entwickelte Know-how für zukünftige Qualifizierungsaktivitäten dauerhaft produktiv gemacht werden konnte. Grundlage der Zusammenarbeit war ein zwischen Hochschulen mit Studienangeboten aus dem Bereich Bibliothek, Archiv, Museum in Deutschland, der Schweiz und Österreich sowie der SUB Göttingen im Jahr 2007 vereinbartes „Memorandum of Understanding“ (MoU 2007). Im Jahre 2011 konnte der Teilnehmerkreis dieser internationalen Ver-

¹² Vgl. – analog zu den Bearbeitungsangaben in Fußnote 11 – den Abschnitt 19.2 in Scheffel/Oßwald/Neuroth (2010).

einbarung sehr unterschiedlicher Hochschulen im Rahmen einer Neuauflage des MoU (MoU 2011) sogar erweitert werden.

Mit diesem MoU ist die Voraussetzung geschaffen worden, nicht nur die nachfolgend skizzierten, kooperativen Publikations-, Projekt- und Qualifizierungsaktivitäten zu realisieren. Der MoU-Kreis ist auch anerkannter Ansprechpartner auf nationaler wie internationaler Ebene, ist durch einen Vertreter im nestor-Steuerungsgremium in die wesentlichen Entwicklungen und Entscheidungen von nestor eingebunden und ermöglicht so, die Kompetenzen und Ressourcen der Hochschulen (bzw. ihrer Mitglieder) gezielt verteilt oder gebündelt in LZA-bezogene Aktivitäten rund um nestor einzubringen.

Hierbei ist die Abteilung F&E der SUB Göttingen nicht nur ein ganz wesentlicher Partner in dieser konzeptionell kooperativ angelegten Zusammenarbeit, sondern durch ihre sonstigen nationalen und internationalen Projektaktivitäten im LZA-Bereich und durch die Rolle der SUB als nestor-Partner auch immer wieder Garant für die Nachhaltigkeit der einzelnen Aktivitäten im Bereich LZA-Qualifizierung.

Langzeitarchivierung digitaler Objekte als Gegenstand von Aus-, Fort- und Weiterbildungsangeboten

Im Kontext der genannten Rahmenbedingungen sind seit 2005 von der Abteilung F&E der SUB Göttingen in Zusammenarbeit mit Partnern aus dem nestor-Umfeld, mit internationalen Partnern aus dem LZA-Bereich sowie insbesondere mit den im MoU zusammengeschlossenen Hochschulpartnern zahlreiche Qualifizierungsangebote im In- und Ausland realisiert worden. Dabei wurden nicht nur traditionelle Vermittlungswege beschränkt (wie Seminare, Präsentationen, Vorträge und Publikationen), sondern auch innovative Verfahren realisiert: z.B. kommentierbare, versionierte Publikationen als sog. „living documents“ (*nestor Handbuch*), zwischen Hochschulen kooperativ konzipierte, in der Realisierung abgestimmte und gegenseitiger Qualitätskontrolle unterworfenen E-Learning-Module oder auch die nestor Schools, bei denen sonst getrennt angesprochene Zielgruppen zusammen und voneinander lernen, LZA-relevante Fragestellungen bearbeiten und dabei möglichst kreativ und hierarchiefrei miteinander kommunizieren (vgl. für die genannten Angebote <http://nestor.sub.uni-goettingen.de/education/index.php>).

Nachstehend erfolgt ein kurzer Überblick zu diesen verschiedenen Angebots- und Vermittlungsformen, die von Scheffel/Oßwald/Neuroth (2010) schon ausführlicher dargestellt wurden.¹³

Als präsentationsbasierte Veranstaltungen konnten die ersten Qualifizierungsangebote einen zahlenbezogen nur begrenzten Teilnehmerkreis von allgemein Interessierten erreichen. Eine erste Maßnahme waren Video-Aufnahmen dieser Veranstaltungen, um diejenigen, die als fachliche Experten ohnehin schon stark in laufenden Projekten (insb. kopal/koLibRI-Entwicklung) involviert waren, nicht noch stärker zu belasten. Diese Aufnahmen wurden dann als unentgeltliche DVDs an Interessierte u.a. bei weiteren nestor-Veranstaltungen verteilt. Die eher auf einführende oder allgemeine Information von grundsätzlich interessierten Personen ausgerichteten (Inhouse-) Seminare wurden ergänzt um vertiefende, zeitlich und inhaltlich intensive Qualifizierungsmöglichkeiten im Rahmen der nestor Schools, die seit 2007 bislang insgesamt sieben Mal¹⁴ an abgeschieden gelegenen Orten stattfanden. Ziel dieser fünf Tage, seit 2011 nur noch drei Tage währenden School-Events war und ist es, ungestört von Kommunikation nach außen in kurzer Zeit einen intensiven, vertieften Einblick in ausgewählte Themen der LZA zu gewinnen. Hierzu wurde ein Konzept realisiert, in dem nach einführenden Fachvorträgen von Experten durch vertiefende Gruppenübungen zuerst eine Reflexion der Fragestellungen in der Teilgruppe und danach eine weiterführende Diskussion im Plenum der Gesamtgruppe erfolgt. Die Perspektiven der verschiedenen Teilnehmer(gruppen) aus dem privatwirtschaftlichen, dem öffentlich-rechtlichen sowie dem Hochschulbereich (Studierende und Professoren) werden dabei möglichst hierarchiefrei ausgetauscht. Die Intensität der Lernatmosphäre wird zudem durch das Zusammensein in der Seminarstätte gesteigert. Die z.T. euphorischen Berichte (vgl. z.B. Häberli/Rösch 2007) über die Lernerfahrungen bei den Schools haben viele Teilnehmer dazu motiviert, wiederholt an diesem Typ von Fortbildungsveranstaltung teilzunehmen.

13 Für weitere Details vgl. auch die Vorversionen des Beitrags von Scheffel/Oßwald/Neuroth (2010) im *nestor Handbuch* (Kapitel 19; <http://nestor.sub.uni-goettingen.de/handbuch/artikel.php?id=75>), auf die hier gemäß dem in Fußnote 11 genannten Verfahren Bezug genommen wird.

14 Die bislang sieben School-Events fanden statt: 2007 (zwei Mal in Gernrode), 2008 (Staufen/Breisgau), 2009 (Staufen/Breisgau), 2010 (Staufen/Breisgau), 2011 (Gernrode) sowie 2012 (Reinhausen b. Göttingen). Alle School-Events sind auf der Website <http://nestor.sub.uni-goettingen.de/education/index.php> ausführlich dokumentiert.

Präsentationen und Vorträge bei nationalen und internationalen Fachveranstaltungen waren und sind ein weiterer, wichtiger Beitrag dazu, die Qualifizierungsaktivitäten transparent zu machen und über innovative Vermittlungsformen zu berichten. Zudem können sie zum generellen Anliegen beitragen, die Sensibilisierung für das Thema Langzeitarchivierung und Langzeitverfügbarkeit zu erhöhen. Auf diese Weise motivieren sie, sich intensiver mit den entsprechenden Fragestellungen zu beschäftigen. Präsentationen z.B. im Rahmen von nestor-Sessions bei den Bibliothekartagen 2008 und 2009,¹⁵ durch Publikationen (vgl. z.B. Neuroth/Oßwald 2008), Vorträge bei internationalen Fachveranstaltungen wie iPRES 2007 (Oßwald 2007) bzw. 2011 (Oßwald/Neuroth/Strathmann 2011) oder DigCCurr 2009 (Neuroth/Oßwald/Strathmann 2009) trugen zur internationalen Wahrnehmung der Qualifizierungsaktivitäten bei und führten umgekehrt auch zu ihrer Unterstützung¹⁶ und schließlich auch zur Einbeziehung dieser speziellen Expertise im LZA-Qualifizierungsbereich in EU-geförderte Projekte wie DigCurV (s.u.).

Die von Anfang an angestrebte und realisierte internationale Vernetzung der Qualifizierungsaktivitäten bezog sich dabei nicht nur auf den deutschen Bereich – wobei die gängige hochschultypbezogene Differenzierung überwunden wurde. Durch die Mitarbeit von Hochschulen aus Österreich und der Schweiz im MoU-Kreis wurden einerseits die gängigen landesbezogenen Grenzen überschritten. Die Zusammenarbeit mit Projekten im EU- bzw. anglo-amerikanischen Bereich und einzelnen Hochschulen aus diesem Kontext ermöglichte andererseits die internationale Orientierung der nestor-Qualifizierungsaktivitäten.

Einen neuen Weg sind die im MoU zusammengeschlossenen Hochschulpartner bei der Erstellung von auf Gegenseitigkeit nutzbaren Lehr- und Lernmaterialien gegangen, mit denen sie auf die jeweils sehr begrenzten personellen und sonstigen Ressourcen an den einzelnen Hochschulen reagierten. Faktisch waren diese Moodle-basierten E-Learning-Module der erste Kristallisationspunkt der Hochschulzusammenarbeit, bei der studentische Seminargruppen aus Chur, Köln, Leipzig, Potsdam und später auch Darmstadt sowie Stuttgart insgesamt 18 E-Learning-Module kooperativ und kollaborativ erarbeiteten. Während einzelne Hochschulen mehrere Module erstell-

15 z. B. bei den Bibliothekartagen 2008 in Mannheim (Oßwald 2008) oder Erfurt (Oßwald/Strathmann 2009)

16 Die nestor Schools 2007 bis 2009 wurden von Projekt „digital preservation europe“ finanziell unterstützt (<http://www.digitalpreservationeurope.eu>).

ten, fokussierten sich andere auf die Aktualisierung dieser in den Hochschulen z.T. regelmäßig in die Curricula einbezogenen Lehr- und Lernmodule. Inwieweit dieses Konzept auch zukünftig noch weitergeführt werden kann, wird im MoU-Kreis beraten.¹⁷

Die Lehre zu LZA-Themen an den einzelnen Hochschulen hat sich im hier betrachteten Zeitraum deutlich intensiviert und professionalisiert. Dies ist nicht zuletzt auch ermöglicht worden durch Grundlagen-Publikationen wie das von Autoren und Herausgebern aus dem MoU-Kreis getragene *nestor Handbuch*¹⁸ oder die 2012 erschienene Bestandsaufnahme zum Thema Forschungsdaten (Neuroth et al. 2012), aber auch durch zahlreiche Fachpublikationen im nestor-Kontext, Qualifizierungsarbeiten und durch nestor-Arbeitsgruppen erstellte Ratgeber und Handlungsanleitungen (vgl. <http://www.langzeitarchivierung.de> ⇒ Publikationen). Hier kann heute – anders als 8–10 Jahre zuvor – auf völlig veränderter, faktisch deutlich verbesserter Grundlage agiert werden. Zahlreiche dieser Veröffentlichungen wurden von der Abteilung F&E initiiert oder sind unter Beteiligung ihrer (Projekt-) MitarbeiterInnen in Kooperation mit anderen Partnern zustande gekommen.¹⁹

Einfluss der Aktivitäten auf die curricularen Angebote

Die Einbindung LZA-spezifischer Themen in curriculare Angebote ist ein über die Jahre immer wichtiger gewordener Bestandteil der Qualifizierungsbemühungen. Zu den wesentlichen Zielen des nestor-MoU gehört einerseits der Ausbau des curricularen Anteils des Themas digitale Langzeitarchivierung und andererseits die kooperative Entwicklung curriculärer Bausteine. Neben den bereits beschriebenen Aktivitäten wurden auch weitere Initiativen unternommen und unterstützt.

17 Die Themen der E-Learning-Module reichen von einer häufig genutzten „Einführung in die Langzeitarchivierung digitaler Objekte“ über OAIS, Metadaten bis hin zu speziellen Anwendungssegmenten wie z.B. Webarchivierung, CAD-Daten oder digitale Fotografien.

18 Vgl. die aktuelle Online-Version 2.3 (Neuroth et al. 2010) bzw. die Printfassung 2.0 (Neuroth et al. 2009).

19 Eine vertiefte Analyse dieses Einflusses könnte Grundlage einer interessanten Studie über den Nachhaltigkeitseinfluss von Projekten und forschungsorientierten Strukturen ergeben.

So hat das Institut für Bibliotheks- und Informationswissenschaft (IBI) der Humboldt-Universität zu Berlin,²⁰ ein Partner im nestor-MoU-Kreis, eine Initiative der EU-Kommission aufgegriffen und bietet, in Kooperation mit dem Kings College London,²¹ einen Masterstudiengang Digital Information & Asset Management²² an. Dieser Ansatz gemeinsamer, kooperativer Studiengänge wird vom MoU-Kreis als eine Möglichkeit erachtet, als notwendig erkannte Bildungsangebote zu realisieren, die einzelne Institutionen nicht oder nicht jetzt anbieten können.

Die Mitarbeit der SUB Göttingen im Projekt DigCurV (vgl. oben) wird durch die anderen Partner des nestor-MoU-Kreises unterstützt, die als assoziierte Partner durch verschiedene Aktivitäten der Mitarbeit zum Gelingen des Projektes beitragen. Ziel des Projektes ist die Entwicklung eines Rahmencurriculums für die berufliche Weiterbildung im Bereich der digitalen LZA. Ein solches Rahmencurriculum erleichtert perspektivisch die Entwicklung von bedarfsgerechten Angeboten zur Weiterbildung im Bereich der digitalen LZA.

Im Einzelfall konnten auch Mitarbeiter der Abteilung F&E der SUB gewonnen werden, um Ihre Erfahrungen im Bereich der LZA in die Hochschulen zu tragen. So konnten verschiedentlich Gastvorträge, Lehraufträge und Dozentenvertretungen realisiert werden. Dies war durch die Nähe der SUB-Mitarbeiter zur Praxis einerseits und zu aktuellen Forschungsfragen andererseits bereichernd für die Hochschulen und lehrreich für die F&E-Vertreter, die sich so ein sehr realistisches Verständnis von den Bedürfnissen der Studierenden erschließen konnten.

Zusammenfassung und Schlussfolgerungen

Die enge Kooperation mit den Hochschulpartnern ist für die Abteilung F&E der SUB sehr bereichernd. Die Erfahrungen, wie das erarbeitete Know-how aus dem Projektalltag in die Lehre transferiert werden kann und muss, wirken auf vielfältige Weise auf die projektbezogenen Arbeiten der Abteilung zurück.

20 Vgl. <http://www.ibi.hu-berlin.de/>.

21 Vgl. <http://www.kcl.ac.uk/index.aspx>.

22 Vgl. <http://www.kcl.ac.uk/prospectus/graduate/index/name/digital-information-and-asset-management>.

Die Erfahrung, erarbeitetes Wissen auch weitergeben und vermitteln zu können, hat dazu geführt, dass in immer mehr Projekten Aufgaben der Qualifikation und Wissensvermittlung mitgedacht werden, dass entsprechende Arbeitspakete bearbeitet werden und dass sich die SUB – auch abseits des hier beschriebenen Gebietes der LZA – verstärkt an solchen Aktivitäten beteiligt.

Ein sehr erfreulicher Nebeneffekt des Engagements für die Aus-, Fort-, und Weiterbildung ist, dass einerseits das eigene Personal beständig fortgebildet wird und andererseits potenzielle künftige Mitarbeiter und Kollegen direkt angesprochen und an aktuelle Aufgaben herangeführt werden können. Im Rahmen der Qualifizierungsaktivitäten ergeben sich immer wieder Kontakte, die – zu beiderseitigem Nutzen – zur Rekrutierung geeigneter neuer Mitarbeiter führen. Zudem ergänzt die enge Kooperation mit den Hochschulpartnern in der Lehre vortrefflich die Ausrichtung der Abteilung F&E auf die Zusammenarbeit mit der Wissenschaft und den Wissenschaftlern zum gemeinsamen Aufbau von Virtuellen Forschungsumgebungen und Infrastrukturen für wissenschaftliche Daten und Dienste.

Aus der Sicht der nestor-MoU-Partner, die sich über die Jahre aus personellen und organisatorischen Gründen mit unterschiedlicher Intensität in die genannten kooperativen Qualifizierungsmaßnahmen einbringen konnten, hat sich die Zusammenarbeit mit der Abteilung F&E der SUB Göttingen als Glücksfall erwiesen. Die organisatorische und im Einzelfall auch personelle Kontinuität hat eine vielgestaltige Vernetzung und Unterstützung ermöglicht. Hierzu hat nicht zuletzt die von allen Partnern anerkannte Professionalität und inhaltliche Kompetenz der KollegInnen in der SUB Göttingen beigetragen. Ohne sie wäre es sicher schwierig gewesen, die durch latente Konkurrenz begleitete Zusammenarbeit der Hochschulen über eine derart lange Zeitspanne aufrecht zu erhalten. Gleichzeitig war die im nationalen und internationalen Wissenschafts- und Förderkontext erfahrene und etablierte F&E-Abteilung der SUB Göttingen für die Hochschulpartner der Schlüssel zu finanzieller und struktureller Unterstützung von wünschenswerten und innovativen Qualifizierungsaktivitäten, die ohne einen solchen Partner nicht zugänglich gewesen wäre.

Aus hochschulpolitischer Sicht ist diese Zusammenarbeit darüber hinaus erneut ein Indiz für die These, dass bei entsprechender Sachorientierung und Kooperationsbereitschaft der beteiligten Personen die Unterschiedlichkeit von Hochschulen und Hochschultypen als synergetische Verstärkung bei der

gemeinsamen Zielerreichung genutzt werden kann. Man muss es eben nur wollen!

Literaturverzeichnis

- Häberli, Justine; Rösch, Bernhard (2007): nestor Spring School 2007 – Einführung in die digitale Langzeitarchivierung in Gernrode/Harz vom 11.–15. März 2007. In: *Bibliotheksdiens*t 41 (2007), H. 5, S. 557–563.
- Memorandum of Understanding (2007): Kooperative Entwicklung curriculärer Module zur digitalen Langzeitarchivierung im Rahmen des nestor II Arbeitspaketes 5. <http://nestor.sub.uni-goettingen.de/education/mou.pdf>.
- Memorandum of Understanding (2011): Kooperative Entwicklung curriculärer Module zur digitalen Langzeitarchivierung. <http://nestor.sub.uni-goettingen.de/education/mou2011.pdf>.
- nestor – Kompetenznetzwerk Langzeitarchivierung (2006): Memorandum zur Langzeitverfügbarkeit digitaler Informationen in Deutschland. <http://files.d-nb.de/nestor/memorandum/memo2006.pdf>.
- Neuroth, Heike et al. (Hrsg.) (2009): *nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.0*. Boizenburg: Hülsbusch.
- Neuroth, Heike et al. (Hrsg.) (2010): *nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.3*. <http://nestor.sub.uni-goettingen.de/handbuch/index.php>.
- Neuroth, Heike et al. (Hrsg.) (2012): *Langzeitarchivierung von Forschungsdaten: Eine Bestandsaufnahme*. Boizenburg: Hülsbusch.
- Neuroth, Heike; Oßwald, Achim (2008): Curriculare Innovation im Spezialbereich: Qualifizierung im Themenbereich „Langzeitarchivierung digitaler Objekte“. In: *ZfBB* 2008: 3, 190–197.
- Neuroth, Heike; Oßwald, Achim; Strathmann, Stefan (2009): Qualification & Education in Digital Curation: the nestor Experience in Germany. In: Helen R. Tibbo et al. (Hrsg.): *Digital Curation: Practice, Promise and Prospect*, Proceedings of DigCCurr2009, April 1–3, 2009, Chapel Hill. Chapel Hill, NC: School of Information and Library Science, University of North Carolina, S. 12–18.
- Oßwald, Achim (2007): nestor-Training & Education: Experiences from Germany. Vortrag bei der iPRES 2007 am 25.10.2007, Beijing. Vortragsfolien unter http://ipres.las.ac.cn/pdf/Achim%20Osswald_iPRES07_111007.pdf.
- Oßwald, Achim (2008): nestor und die Aus- und Weiterbildungsangebote zur Langzeitarchivierung. Vortrag beim Bibliothekartag 2008 in Mannheim. Vortragsfolien unter <http://www.opus-bayern.de/bib-info/volltexte/2008/534/>.

- Oßwald, Achim; Neuroth, Heike; Strathmann, Stefan (2011): Curation and Preservation of Research Data in Germany. A survey across different academic disciplines. In: Borbinha, J. et al. (Hrsg.): *iPRES 2011 – 8th International Conference on Preservation of Digital Objects*. Proceedings. Lisbon: Biblioteca Nacional de Portugal, S. 274–275.
- Oßwald, Achim; Scheffel, Regine (2006): Lernen und weitergeben – Aus- und Weiterbildungsangebote zur Langzeitarchivierung. Folien des Vortrags bei „3 Jahre nestor – Abschlussveranstaltung“, Frankfurt, 19.6.2006. http://files.d-nb.de/nestor/veranstaltungen/2006-06_19/nestor_2006_06_19_osswald_scheffel.pdf.
- Oßwald, Achim; Scheffel, Regine (2007): Lernen und Weitergeben – Aus- und Weiterbildungsangebote zur Langzeitarchivierung. In: Neuroth, Heike et al. (Hrsg.): *nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 0.1*. Göttingen: nestor – Kompetenznetzwerk Langzeitarchivierung digitaler Ressourcen für Deutschland; S. 119–126. http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_22.pdf.
- Oßwald, Achim; Strathmann, Stefan (2009): nestor Weiterbildung: Tutorials, Handbuch und Schools. Vortrag beim Bibliothekartag 2009 in Erfurt. Vortragsfolien unter http://www.opus-bayern.de/bib-info/frontdoor.php?source_opus=752.
- Scheffel, Regine; Oßwald, Achim; Neuroth, Heike (2010): Qualifizierung im Themenbereich „Langzeitarchivierung digitaler Objekte“. In: Neuroth, Heike et al. (Hrsg.): *nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 2.3*. Göttingen: nestor – Kompetenznetzwerk Langzeitarchivierung digitaler Ressourcen für Deutschland; S. 19:1–19:20. http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_383.pdf.

Digital Humanities lehren und lernen

Modelle, Strategien, Erwartungen

von Malte Rehbein und Patrick Sahle

Einleitung

Auch wenn die Digital Humanities (DH) unter ihren wechselnden Bezeichnungen¹ seit mindestens 60 Jahren existieren und seit langer Zeit eine eigene Infrastruktur aus Verbänden, Zeitschriften, Mailing-Listen, Forschungsprojekten und akademischen Einrichtungen entfalten, scheint dieses akademische Feld in den letzten Jahren nochmals besonders an Schwung gewonnen zu haben.² Hier ist besonders eine Zunahme der Forschungsprojekte – auch dank spezieller Förderlinien der Drittmittelgeber – zu beobachten, die zu einer steigenden Bedeutung und besserer Sichtbarkeit digitaler Forschung in den etablierten Fächern und der akademischen Landschaft insgesamt beitragen.

Während die DH einerseits als Transformationsprozess der Geisteswissenschaften verstanden werden können, bilden sie andererseits einen eigenen Forschungsbereich und zunehmend auch ein eigenes Lehrfach. Gerade für die andauernde Etablierung der DH scheint es von besonderer Bedeutung zu sein, dass sie auch in Studienprogrammen vermittelt werden, wobei sich diese Lehre auf verschiedene Ebenen der Ausbildung beziehen kann und damit auch verschiedene Zielgruppen adressiert: Summer Schools und Workshops vermitteln Kenntnisse vor allem an bereits ausgebildete Wissenschaftler; Module in Ph.D.-Programmen und Doktorandenschulen ergänzen Fachstudien und stellen ein neues methodisch-technisches Rüstzeug bereit; B.A.- und M.A.-Studiengänge erlauben eine fachliche Spezialisierung schon wäh-

1 Der letzte Namenswechsel, vom „Humanities Computing“ zu den „Digital Humanities“, liegt noch nicht einmal eine Dekade zurück.

2 Jüngstes Zeugnis dieser Entwicklung ist etwa die „Global Outlook :: Digital Humanities“-Initiative von ADHO (s. <http://www.globaloutlookdh.org>). Siehe auch Gold (2012: ix–xvi).

rend des Studiums; einzelne Kurse, Module und abgestimmte Zertifikate ermöglichen eine zusätzliche Profilbildung in den gängigen geisteswissenschaftlichen Ausbildungen. Die anhaltende Ausweitung der DH-Lehre erfordert eine internationale Abstimmung über das Selbstverständnis des Faches, über seine Inhalte, die unterschiedlichen pädagogisch-didaktischen Ansätze und schließlich möglichst konvergente curriculare Entwicklungen und gemeinsame Zielstellungen.³

Der vorliegende Beitrag versucht, einen ersten Einblick in bestehende, typische Studienprogramme, ihre thematische Ausgestaltung und die möglichen Abschlüsse zu geben. Er thematisiert dann vor allem die Beziehungen zwischen den DH als Fach, ihren Lehrinhalten, Lehrformen und den Zielkompetenzen von Absolventen – und wie sich diese in Berufsperspektiven auf dem Arbeitsmarkt niederschlagen.

Studiengänge am Beispiel

Weltweit gibt es inzwischen eine große Zahl von Studiengängen im Bereich der Digital Humanities. Wie groß diese Zahl genau ist, hängt wesentlich davon ab, was man als DH-Degree definieren will: Sie liegt dann im Jahr 2012 zwischen 15 bei einer engen und knapp 100 bei einer weiten Definition.⁴ Paradigmatisch sind aber zweifellos vor allem jene Programme, die sich selbst als „Digital Humanities“ (oder analog in anderer nationalsprachlicher Übersetzung)⁵ bezeichnen und den Anspruch erheben, die DH in ihrer gan-

3 Dementsprechend entsteht nun auch sukzessive Literatur zur Lehre in den DH. Siehe zuletzt vor allem Hirsch (2012); auch Gold (2012) bietet mit Sektion V vier Beiträge zur Lehre; der noch immer als Standardwerk zu DH angesehen „Companion“ von 2004 hat noch keinen dezidierten Abschnitt zu Fragen der Didaktik oder der Ausbildungspolitik; einzelne Abschnitte verweisen jedoch auf die Verknüpfung von Forschung und Lehre in den DH (vgl. Schreibman/Siemens/Unsworth 2004).

4 Einen Beitrag zur Definition und Differenzierung von DH-Degrees wird in Kürze auch ein Bericht über die Ausbildungssituation in den DH für DARIAH-DE (<http://de.dariah.eu/>) liefern.

5 Im Deutschen gibt es dazu keinen eingeführten Begriff. Allein die Kölner „Informationsverarbeitung [in den Geisteswissenschaften]“ könnte als traditionelles Gegenstück verstanden werden. Die Formulierung „Digitale Geisteswissenschaften“ als direkte Entsprechung hat hier keine Tradition. Im Italienischen ist „informatica umanistica“ seit langer Zeit gut eingeführt, im Französischen „humanités numériques“, eine eher neue Schöpfung, ebenso wie im Spanischen „humanidades digitales“.

zen fachlichen Breite und methodisch-technischen Tiefe abzudecken. Dazu werden hier zunächst vier besonders etablierte und damit in gewissem Sinne auch prototypische Studiengänge beschrieben – auch wenn in diese Gruppe noch eine Reihe weiterer Programme fallen würde.⁶

Beispiel 1: Das Fach „Informationsverarbeitung“ an der Philosophischen Fakultät der Universität zu Köln ist ein konsekutiver Studiengang, der B.A. und M.A umfasst. Eine Promotion ist ebenfalls möglich. B.A. und M.A. sind Zwei-Fach-Studiengänge und umfassen, wenn sie als Hauptfach studiert werden, 96 bzw. 82 ECTS. Das Fach besteht seit 1997 (ursprünglich als Magisterstudiengang), ist seit vielen Jahren sehr stark nachgefragt und häufig zulassungsbeschränkt. Regelmäßig beginnen ca. 80 Studierende im Wintersemester mit dem B.A.-Studium, allerdings sind hier auch die Studierenden der „Medieninformatik“ (einem Teilfach der „Medienwissenschaften“) mit eingerechnet, deren Kurse dem Angebot der Informationsverarbeitung entnommen sind. Das Kölner Angebot wird von zwei Lehrstühlen getragen, die bereits Anfang der 1990er-Jahre eingerichtet wurden und vom Zuschnitt her die Geisteswissenschaften in ihrer ganzen Breite abdecken sollen: Die Professur für „Sprachliche Informationsverarbeitung“ zielt vor allem auf Computerlinguistik und Texttechnologie, während die Professur für „Historisch-Kulturwissenschaftliche Informationsverarbeitung“ die übrigen geisteswissenschaftlichen Problemstellungen im Blick hat und hierbei neben textlichen vor allem auch bildliche oder objektgebundene Informationen bearbeitet.⁷ Beide verstehen sich ausdrücklich als „geisteswissenschaftliche Fachinformatik“ und betonen die Bedeutung nicht nur theoretischer Arbeiten in den DH, sondern auch des Software-Engineering auf der Basis höherer Programmiersprachen (Java oder C++). Absolventen sollen dadurch in der Lage sein, geisteswissenschaftliche Fragestellungen nicht nur zu analysieren, für for-

6 Zu nennen sind hier – um nur den Master-Bereich einigermaßen abzudecken – noch der „MA/MSc in Digital Humanities“ am University College London, der „M.A. in Digital Humanities and Culture“ am Trinity College Dublin, der „M.A. in Digital Humanities“ an der National University of Ireland in Maynooth, der „Master in Digital Arts and the Humanities“ am University College Cork, der „MA in Digital Humanities“ an der Loyola University Chicago sowie die „Laurea Magistrale in Informatica Umanistica“ an der Università di Pisa.

7 Für die „Sprachliche Informationsverarbeitung“ siehe <http://www.spinfo.phil-fak.uni-koeln.de/>, für die „Historisch-Kulturwissenschaftliche Informationsverarbeitung“ <http://www.hki.uni-koeln.de/>.

male Methoden zugänglich zu machen und bestehende Software anzupassen und einzusetzen, sondern selbst neue Methoden und Softwarelösungen zu entwickeln.

Beispiel 2: Die Universität Würzburg hat als Nachfolger des Aufbaustudienganges „EDV-Philologie“ im Wintersemester 2009/10 einen B.A. „Digital Humanities“ eingeführt, der zum Wintersemester 2012/13 um einen konsekutiven M.A. mit demselben Namen erweitert worden ist. Federführend für beide Würzburger Studiengänge ist der Lehrstuhl für „Computerphilologie und Neuere Deutsche Literaturgeschichte“ an der Philosophischen Fakultät I, an der auch die Promotion möglich ist. Der B.A. kann sowohl als Hauptfach (85 ECTS) als auch als Nebenfach (60 ECTS) studiert werden. Der Studiengang ist zulassungsfrei und zieht pro Jahr zwischen 50 und 60 Studierende an, die die Digital Humanities immer in Kombination mit einem anderen (frei wählbaren) kultur- oder geisteswissenschaftlichen Fach studieren. Digital Humanities in Würzburg werden damit als ein „interdisziplinär ausgerichtetes Fach“ verstanden, „dessen Vertreter sowohl durch eine traditionelle Ausbildung in den Geistes- und Kulturwissenschaften ausgewiesen sind wie durch ihre Vertrautheit mit Konzepten, Verfahren und Standards der Informatik“⁸. Dies zeigt sich im Aufbau des Studiums, das neben den eigentlichen DH-Kernveranstaltungen auch den Besuch geisteswissenschaftlicher „Grundkurse“ (in anderen Fächern als dem Zweitfach) sowie eine Einführungsvorlesung Informatik an der Fakultät für Mathematik und Informatik vorsieht. Trotz der formalen Verankerung des Studienganges am Institut für deutsche Philologie werden Lehrveranstaltungen der DH auch in Kooperation unter anderem mit der Geschichtswissenschaft und der klassischen Philologie angeboten und von Dozenten dieser Disziplinen geleitet. Im Vordergrund des Würzburger Studienganges steht die Methodenkompetenz, insbesondere die Modellierung geistes- und kulturwissenschaftlicher Aufgabenstellungen, für die konkrete Techniken und Verfahren lediglich als Beispiele dienen sollen. Typische Studieninhalte sind: Digital Humanities im Überblick, Datenmodellierung, Textkodierung, Gestaltung, Informatik, Programmieren, Digitale Bibliotheken und Informationssysteme sowie eine Projektarbeit. Ein inhaltlicher Schwerpunkt wird auf den Bereich „Edition und Textkorpora“ gelegt.

⁸ Selbstbeschreibung des Studienganges: <http://www.germanistik.uniuerzburg.de/lehrstuehle/computerphilologie/studium/>

Beispiel 3: Am King's College in London besteht schon seit langer Zeit mit dem „Department of Digital Humanities“ (DDH; früher: „Centre for Computing in the Humanities“ – CCH) das wohl größte und renommierteste DH-Zentrum weltweit. Zuletzt mit drei Professuren und 17 Mitarbeitern aus dem „academic staff“ sowie 23 weiteren Projektmitarbeitern ausgestattet, werden drei Master-Programme und ein PhD-Programm aus dem Bereich der Digital Humanities angeboten. Während der „Master in Digital Culture and Society“ bzw. der „Master in Digital Asset Management“ eine eher spezialisierte Perspektive ermöglichen, steht der seit 2002 laufende „Master in Digital Humanities“⁹ paradigmatisch für den ganzen Forschungsbereich. Als Ziele werden hier vor allem die Fähigkeit genannt, ein kritisches Verständnis für den Einsatz digitaler Technologien in den Geisteswissenschaften zu entwickeln sowie praktische Kompetenzen der Informatik aufzubauen, mit denen digitale Ressourcen kreiert werden können. Dem angelsächsischen Verständnis entsprechend, nach dem spezialisierte Master- auf eher allgemeine Bachelor-Studiengänge folgen, gab es bislang keinen DH-B.A. Ab 2014 soll allerdings ein B.A. in „Digital Culture“ angeboten werden. Der „M.A. in DH“ als Ein-Fach-Studiengang erfordert 90 ECTS, kann in einem Jahr Vollzeit oder zwei Jahren Teilzeit absolviert werden, ist (für britische Staatsbürger) mit einer Gebühr von 7500 £ belegt und hatte im Studienjahr 2012/2013 fünfzehn eingeschriebene Studierende.

Beispiel 4: Die kanadische University of Alberta in Edmonton bietet in ihrem Graduate Program Abschlüsse in „Humanities Computing“ entweder als zweijährigen M.A. (Master of Arts, 27 credit hours¹⁰) oder als dreijährigen „combined degree“ M.A./M.L.I.S. (Master of Library and Information Studies) an.¹¹ Der zweijährige M.A. sieht eine Spezialisierung durch Kombination mit anderen kulturwissenschaftlichen Disziplinen, die am Humanities-Computing-Programm mitwirken, oder durch einen individuellen Studienplan vor. In der M.A./M.L.I.S.-Kombination bringen die Studenten gleichwertig Lehrveranstaltungen aus beiden Bereichen, Humanities Computing und Library and Information Studies, ein. Die Programme der University of

9 Zum M.A. in „Digital Humanities“ am King's College London s. <http://www.kcl.ac.uk/artshums/depts/ddh/study/pgt/madh/index.aspx>.

10 neun Veranstaltungen, die jeweils dreistündig unterrichtet werden

11 Zu den Programmen an der University of Alberta s. <http://www.ois.ualberta.ca/en/HumanitiesComputing.aspx>.

Alberta adressieren “the demand for Arts graduates to be able to work either in the realm of humanities research and teaching or in the emerging job markets of information management and content delivery over the internet” und zeigen sich daher vor allem durch den Arbeitsmarkt motiviert. Die Ausbildung zielt auf die Entwicklung technischer Fertigkeiten und sozialer Kompetenzen, um Forschungsprobleme der Humanities entweder selbst oder gemeinsam mit Computerspezialisten zu lösen. Die Lehrveranstaltungen umfassen sowohl die theoretische Annäherung an solche Forschungsprobleme, als auch technologische Konzepte und Praktiken sowie Projektdesign und -management. Der zweijährige M.A. kostet für Inländer rund 6.000 kanadische Dollar und zieht jährlich zwischen fünf und zehn Studierende an. Er ist nicht konsekutiv, sondern schließt an die generelle Liberal-Arts-Ausbildung der Undergraduates an.

Die weitere Landschaft der DH-Lehre

Jenseits der prototypischen Studiengänge wird die Landschaft der DH-Ausbildung diffus. Hier findet man unter allen möglichen Bezeichnungen sehr unterschiedliche Angebote, die sich aber selbst durchaus den Digital Humanities zurechnen würden. Dies können zunächst traditionelle Fächer in Hinwendung zu digitalen Verfahren sein: Seit Langem etabliert sind die Computerlinguistik und die Computerphilologie, neuerdings zeigen aber auch „Digital History“ oder „Archäoinformatik“ ähnliche Tendenzen. Häufiger begegnen einem neu geschaffene Fächer, die vor allem auf den Professionalisierungsbereich der Master zielen und dort starke berufsorientierte Spezialisierungen bieten und zugleich Querbezüge zu anderen Fächern aufweisen. Genannt seien hier Programme zum digitalen Kulturerbe (EuroMACHS = „European Heritage, Digital Media and the Information Society“ in Turku/Coimbra/Graz/Köln; „Médiation culturelle, patrimoine et numérique“ in Paris), zu digitalen Dokumenten und Editionen („Patrimoine écrit et édition numérique“ in Tours; „Information, Communication, Culture et Documentation“ in Lille; „Informatica del Testo ed Edizione Elettronica“ in Siena, „Editions- und Dokumentenwissenschaft“ in Wuppertal), zu digitaler Kultur, Gesellschaft und Wissenschaft („Digital Culture and Society“ in London; „Masterprogram i digital kultur“ in Bergen; ITASIA = „Information, Technology, and Society in Asia“ in Tokio) oder zur Verwaltung und Sicherung digitaler Daten („Digital Asset Management“ in London; „Information Ma-

nagement & Preservation [Digital]“ in Glasgow)¹². Während hier die Überschneidung zu den „Digital Humanities“ sehr groß ist, stellt sich jedoch manchmal die Frage, ob es sich noch um „Humanities“ im engeren Sinne handelt. Eine ähnliche Frage stellt sich auch, wenn DH aus der Hinwendung der Informatik oder der Informationswissenschaft zu diesem Bereich entstehen sollen. Eine auf die Geisteswissenschaften angewandte Informatik ist noch keine Geisteswissenschaft. Und auch in den möglichen Kombinationsstudien „geisteswissenschaftliches Fach plus Informatik“ ist fraglich, ob aus der einfachen Summe von Humanities und Computing der spezifische Bereich „Digital Humanities“ wirklich abgedeckt werden kann.

Erweiterungen erfährt das Fach, gerade in der universitären Lehre, durchaus auch in andere Richtungen, etwa durch die Einbeziehung von Visual/Performing Arts, Creative Studies oder Media Studies etc., vor allem in Nordamerika. So zeigt beispielsweise die kanadische Brock University mit „Interactive Arts and Sciences“, angeboten durch das „Centre for Digital Humanities“, eine sehr weite Verwendung des Etiketts DH: „students will explore the interplay of digital media and our understanding of human identity and values with careful attention to ways in which discourses in technoculture, narrative, gameplay, computing and media production intersect“¹³. Hier wird die ursprüngliche Definition, dass Digital Humanities primär Humanities sind und auf computerbasierte Methoden zu geisteswissenschaftlicher Forschung, Lehre (und Publikation?) zielen, fast schon verlassen und so wird zweifelhaft, ob es sich als Fach noch um einen DH-Studiengang handelt.

Im Regelfall steht eine DH-Ausbildung immer noch in Kombination mit einer Ausbildung in einer oder mehreren Geisteswissenschaften oder mit allgemeineren Modulen, die über die geisteswissenschaftliche Einzelwissenschaft hinausgehen. Die Lehrgegenstände in den einzelnen DH-Programmen können dabei sehr unterschiedlich aussehen – je nachdem, welcher fachliche Hintergrund oder spezialisierende Zuschnitt für ein Angebot maßgeblich ist. Trotz aller Unterschiede scheinen sich die Inhalte der DH im engeren Sinne

12 Auf diesem Feld gibt es sogar eigenständige Curricularbemühungen, siehe DigCurV (Digital Curator Vocational Education Europe, <http://www.digcur-education.org/>) für die europäische und DPOE (Digital Preservation Outreach & Education, <http://www.digitalpreservation.gov/education/index.html>) für die amerikanische Seite.

13 <http://www.brocku.ca/webcal/2009/undergrad/digi.html>

und damit auch die Lernziele auf einer abstrakteren Ebene näherungsweise in den folgenden Gruppen fassen zu lassen:

- Kenntnis über Entwicklung, Stand und Fragestellungen der DH insgesamt
- Grundlagen der Informatik und Informationswissenschaft
- Angewandte Informatik (z.B. Webtechnologien, Programmierung)
- technische Standards, Datenstandards, Metadatenstandards
- Schwerpunktthemen der DH (z.B. Digital Libraries, Texts and Corpora, Scholarly Digital Editions)
- Erzeugung und Bewertung digitaler Daten (Primärdaten, Auswertungsdaten) und Medien
- Anwendung digitaler Verfahren auf Fragen der Geisteswissenschaften in einer digitalen Umwelt mittels Modellierung und Formalisierung der Informationsstrukturen und Fragestellungen
- kritische Reflexion des digitalen Wandels von Wissenschaft und Gesellschaft
- kollaborative Arbeitsformen in multidisziplinären Teams und Projekten.

Dies kann nur eine sehr allgemeine Beschreibung sein. Die vielfältigen Programme setzen jeweils eigene Schwerpunkte und decken die genannten Felder selbst intentional sehr unterschiedlich ab. Trotzdem dürften sie so etwas wie ein „gemeinsames Verständnis“ oder einen gemeinsamen Kern des Faches umreißen.

Im Einzelfall und im Detail ist dann z.B. zu sehen, ob eine besondere Orientierung an bestimmten Bereichen der Geisteswissenschaften (wie z.B. den Philologien und Sprachwissenschaften) besteht und wie das Verhältnis von Theorie und Praxis ausgestaltet wird. Hier lassen sich die Programme z.B. nach ihrem Informatikanteil beschreiben: Während manche Studiengänge Informatikgrundlagen und höhere Programmierkenntnisse enthalten, wie sie auch in „normalen“ Informatikstudiengängen anzutreffen sind, beschränken sich andere auf die „einfacheren“ Webtechnologien,¹⁴ spezielle Texttechnologien und jene Datenformate und -standards, die für Objekte der Geisteswissenschaften von besonderer Relevanz sind. Dazu gehören dann auch die

14 Hier reicht das Feld dann von den Client-Server-Architekturen, Protokollen, HTML, CSS und Javascript bis hin zu SQL-Datenbanken, Content-Management-Systemen und Repositories, den „X“-Technologien (auf der Basis von XML) und Scriptsprachen wie PHP oder Python.

verschiedenen generischen und fachspezifischen Ansätze und Standards zur Beschreibung von Metadaten (z.B. DC oder METS), Daten (vor allem TEI) und semantischen Informationen (z.B. RDF oder OWL).

Eine DH-Ausbildung deckt das Feld also in unterschiedlicher Breite ab. Zugleich ergibt sich aber auch eine bestimmte Tiefe dadurch, dass DH im Studienverlauf zu verschiedenen Zeiten und mit einem unterschiedlichen Anteil am individuellen Gesamtprogramm belegt wird. Dies schlägt sich unter anderem in unterschiedlichen Abschlüssen nieder.

Abschlüsse

Im Versuch, DH-Abschlüsse in ihrer ganzen Bandbreite zu systematisieren, spiegeln sich die unterschiedlichen Sichtweisen auf die Rolle, die die DH im Gesamtkontext der Geistes- und Kulturwissenschaften spielen, und die daraus abgeleiteten Ausbildungsziele wider. Diese lassen sich grob in zwei Gruppen einteilen.

In der ersten Gruppe wird vor allem von der Beobachtung ausgegangen, dass zu Beginn des 21. Jahrhunderts neue Medien und Computertechnologie allgegenwärtig sind und immer mehr den Alltag und die Gesellschaft im Allgemeinen und die Arbeit innerhalb der Geistes- und Kulturwissenschaften im Speziellen durchdringen, diese Disziplinen mithin einen Transformationsprozess durchmachen. Fast wichtiger als Medien und Technologien an sich erscheinen hierbei aber die Paradigmenwechsel, die diese nach sich ziehen. Durch die „Medienrevolution“ unserer Zeit verändert sich nicht nur der Zugang zu Daten, Information und Wissen, sondern vor allem der Umgang mit ihnen und die Kommunikation über sie. Die Geistes- und Kulturwissenschaften erfahren hierdurch einen „computational turn“ mit all seinen größtenteils noch unerforschten, aber wohl unvermeidlichen Konsequenzen.¹⁵ Hier sind die DH gefordert, künftigen Absolventen den kritisch-reflektierenden Umgang in und mit der digitalen Welt zu vermitteln,¹⁶ wie es z.B. die University of Virginia mit ihrem „Praxis Program“ umzusetzen versucht.¹⁷

15 am Beispiel der Geschichtswissenschaft jüngst beschrieben bei Haber (2011)

16 Das Englische kennt hierzu auch den Begriff der „digital literacy“; vgl. Reid (2012: 357 ff.). Die Gesellschaft für Informatik (GI) fordert das Gegenstück zu einer „digital literacy“ aus der Sicht der Informatik mit ihrer Mitteilung „Auswirkungen der Informatik auf die Gesellschaft erforschen und diskutieren“ (s. <http://www.gi.de/aktuel->

Lehrangebote hierzu, die deutlich über die Anwendung von Word, Excel und Powerpoint hinausgehen, werden essenzieller Bestandteil einer jeden geisteswissenschaftlichen Ausbildung – insbesondere auch des Lehramts – sein müssen. Innerhalb dieser Gruppe sind Kompetenzen am „unteren“ Ende der DH folglich integraler Bestandteil der klassischen Studiengänge und ihrer Abschlüsse; die DH spielen aus dieser Perspektive, von den geisteswissenschaftlichen Fächern her betrachtet, eher die Rolle einer „Hilfswissenschaft“, eines elementaren methodischen Handwerks. Wie etwa die Paläografie für die historischen Disziplinen, ohne die das Lesen, Verstehen und letztendlich Interpretieren handschriftlicher Quellen nicht möglich wäre, sind die stark anwendungsbezogenen Teile der DH essenzielle Voraussetzung für das Betreiben von Geistes- und Kulturwissenschaften im digitalen Zeitalter. Diese Angebote unterschreiten dann aber die Schwelle „expliziter“ Studiengänge und sind eher als „implizite“ DH-Lehre zu bezeichnen.

Die zweite Gruppe von Lehrangeboten bildet hingegen DH-Spezialisten aus, die befähigt werden, eben genau diese Voraussetzungen methodisch weiterzuentwickeln und praxistauglich zu machen. Es ist hierbei ein zunehmender Bedarf an professionell ausgebildeten Experten der DH, die Schlüsselaufgaben wie etwa Beratung, Analyse, Datenmodellierung oder Projektmanagement in künftigen geisteswissenschaftlichen und kulturnahen Projekten übernehmen können, festzustellen. Weiterhin ist zu beobachten, dass auch diese Projekte dem generellen Paradigmenwechsel durch den „computational turn“ unterliegen. Sie sind zunehmend durch interdisziplinäre, internationale Zusammenarbeit in unterschiedlich zusammengesetzten, oft wechselnden und häufig geografisch verteilten Teams geprägt und erfordern den Umgang mit komplexen Technologien, die kurzen Entwicklungszyklen unterliegen. Aus den für die genannten Schlüsselaufgaben erforderlichen Kompetenzen leiten sich damit für die DH-Ausbildung deutlich über den kog-

les/meldungen/detailansicht/article/auswirkungen-der-informatik-auf-die-gesellschaft-erforschen-und-diskutieren.html).

- 17 “The Praxis Program realigns graduate methodological training with the demands of the humanities in the digital age. Our goal is to equip knowledge workers for faculty positions or alternative academic careers at a moment in which new questions can be asked and new systems built. The Praxis Program produces humanities scholars who are as comfortable writing code as they are managing teams and budgets” (University of Virginia Library, The Praxis Program, s. <http://praxis.scholarslab.org>).

nitiven Bereich hinausgehende affektive Lernziele ab, die unseres Erachtens bisher noch nicht ausreichend adressiert worden sind.

Zusammenfassend lassen sich die DH-Programme gemäß ihrer Ziele also in zwei große Gruppen gliedern: Einerseits geht es um die Vermittlung von methodischem Rüstzeug, die innerhalb der „normalen“ Studiengänge stattfinden kann. Andererseits geht es um die Ausbildung von DH-Spezialisten zur Weiterentwicklung dieses Rüstzeugs und zur Übernahme von Schlüsselfunktionen in künftigen Projekten. Während in der ersten Gruppe für die erfolgte oder erfolgreiche Teilnahme an Lehrveranstaltungen zumeist Credit Points vergeben werden, die auf das Studium etwa als Schlüsselqualifikation angerechnet werden¹⁸ oder zu einer zusätzlichen Zertifizierung führen können,¹⁹ zielt die zweite Gruppe auf einen spezialisierten, expliziten Studienabschluss.

Hier wiederum ist zwischen den zwei Stufen der nach Bologna reformierten Studienabschlüsse (Bachelor und Master) sowie der akademischen Weiterbildung zu unterscheiden. Die nordamerikanischen Studienprogramme, die traditionell nach Undergraduate- und Graduate-Ausbildung unterscheiden, sind hier vergleichbar, schließen aber in die Graduate-Ausbildung neben dem Master in der Regel die (strukturierten) Ph.D.-Programme ein, die hierzulande weniger verbreitet sind. Dagegen sind die durch die Bologna-Reform in Europa vorgesehenen konsekutiven Studiengänge in Nordamerika unüblich, sodass vergleichbare Angebote wie die konsekutiven B.A./M.A.s in DH in Würzburg oder Köln dort noch fehlen. Das nordamerikanische Modell erlaubt während des Undergraduate-Studiums hingegen individuelle Schwerpunktsetzungen, die an verschiedenen Universitäten²⁰ durch das Angebot von

18 Zu nennen vor allem das Hamburger Modell der „embedded Digital Humanities“ (Digital Humanities im deutschsprachigen Raum, s. <http://www.dig-hum.de/studienstandort/hamburg>); aber auch Institutionen, die ein dezidiertes DH-Programm der zweiten Gruppe betreiben, bieten oft auch ausgewählte Kurse für Hörer anderer Fächer an (etwa am King’s College London).

19 Am ausgereiftesten ist hierbei wohl das IT-Zertifikat der Universität zu Köln, das eine übergreifende DH-Ausbildung bzw. Profilbildung für Geisteswissenschaftler aller Fächer bietet (vgl. oben). Auch auf der Graduate-Ebene existieren Angebote, wie etwa das der University of California, Los Angeles (UCLA) (s. <http://www.cdh.ucla.edu/instruction/dhgradcertprog.html>).

20 In der Regel sind dies jedoch Institutionen, die auch eine Graduierten-Ausbildung anbieten. Für die Integration von DH in nordamerikanische Vier-Jahres-Colleges sprechen sich Alexander und Davis (2012) aus.

„Minor“-Programmen in DH ermöglicht werden.²¹ Dabei wird DH jedoch nicht überall als ein integrativer Studiengang verstanden. So fasst die University of Illinois at Urbana-Champaign DH als Belegung eines separaten Informatik-Minors in Kombination mit einem geisteswissenschaftlichen Major auf.²² Verschiedene deutsche Universitäten folgen einem ähnlichen Ansatz (etwa Bamberg und Jena) – die Informatik hat hier aber ein höheres Gewicht, wenn sie auch als echtes Neben- oder sogar Hauptfach studiert werden kann.²³

Explizite Studienangebote im Hauptfach/Major auf der Bachelor-Ebene, wie in Würzburg oder Köln, sind eher die Ausnahme. Auf der Master-Ebene sind die Angebote breiter gestreut. Hier sind an verschiedenen Universitäten ausgewiesene Abschlüsse als MA in Digital Humanities zu finden.²⁴ Die Bezeichnung der Abschlüsse variiert jedoch von Institution zu Institution, um der jeweiligen Situation vor Ort gerecht zu werden oder eine speziellere Profilbildung zu erreichen. Dies ist auch ein Zeichen dafür, dass sich DH flächendeckend bislang als Begriff doch noch weniger durchgesetzt haben als als akademische Disziplin, wobei eine allgemeine Tendenz festzustellen ist, dass auch die begriffliche Durchsetzung allmählich voranschreitet.

Außerhalb der hier versuchten Systematik der DH-Studiengänge stehen die inzwischen vielseitigen Angebote zur akademischen Weiterbildung für (Nachwuchs-) Forscher und Lehrende, etwa – um wieder mit deutschen Beispielen zu beginnen – die European Summer School in Digital Humanities „Culture & Technology“ an der Universität Leipzig,²⁵ die Paderborn-Detmolder Edirom Summer Schools²⁶ oder die institutionsunabhängigen, regelmäßig an verschiedenen Orten veranstalteten Sommerkurse des Instituts für Doku-

21 Exemplarisch sind hier zu nennen: der „Digital Humanities Minor“ von UCLA (s. <http://www.cdh.ucla.edu/instruction/dhminor.html>); die „Spezialisierung in Digital Humanities“ der Michigan State University (s. <http://www.reg.msu.edu/academicprograms/ProgramDetail.asp?Program=5691>).

22 <https://rhino.ncsa.illinois.edu/display/infominor/Home>

23 Zu Bamberg (Angewandte Informatik/Computing in the Humanities) s. <http://www.uni-bamberg.de/ai/studium>, zu Jena (Informatik als Ergänzungsfach) s. <http://www.fmi.uni-jena.de/Studieninteressierte/Studiengänge/Bachelor/Informatik+Ergänzungsfach.html>.

24 Eine umfassende und aktuelle Übersicht über die weltweit existierenden Programme existiert trotz mancher Bemühung noch nicht.

25 http://www.culingtec.uni-leipzig.de/ESU_C_T/

26 <http://www.edirom.de/summerschool>

mentologie und Editorik (IDE)²⁷. Die starke Nachfrage solcher Angebote zeugt von der bereits vorhandenen Durchdringung von Methoden der Digital Humanities in vielen Bereichen der Geistes- und Kulturwissenschaften, etwa der Textedition und der Textanalyse. Eine Weiterbildung in DH scheint dabei nicht nur für die Durchführung vieler Forschungsprojekte notwendig zu sein, sondern auch als eine hilfreiche Zusatzqualifikation für die Entwicklung akademischer Karrieren angesehen zu werden. Dies bezeugt gleichzeitig aber noch einen Mangel an Angeboten innerhalb vieler Universitäten. Mit anderen Worten: DH werden vielerorts betrieben, die Vermittlung der benötigten Kompetenzen erfolgt aber nur punktuell und teilweise sogar autodidaktisch.

Auch international zeigt sich eine starke Nachfrage an Weiterbildungsangeboten, belegbar etwa durch die kontinuierlich steigenden Teilnehmerzahlen des jährlichen „Digital Humanities Summer Institute“ (DHSI) der University of Victoria (<http://www.dhsi.org>) in Kanada.²⁸ Zudem bestehen, initiiert durch die University of Victoria, Bemühungen zur Zusammenführung der verschiedenen Angebote auf internationaler Ebene, u.a. mit dem Ziel, ein „virtuelles Curriculum“ zu DH als Weiterbildung zu etablieren und den (in der Regel graduierten) Teilnehmerinnen und Teilnehmern an mehreren solcher Veranstaltungen die erfolgreiche Teilnahme zu zertifizieren. Initiativen wie etwa die THATCamps (The Humanities and Technology-Camps, <http://thatcamp.org>) oder die von ADHO ins Leben gerufene Arbeitsgruppe „Global Outlook :: Digital Humanities (GO::DH)“ (vgl. O'Donnell 2012) versuchen, diese Bemühungen auch außerhalb der bisherigen Zielgruppen internationaler Zusammenarbeit (v.a. Nordamerika und größere Teile Europas) zu tragen.

Programme auf den verschiedenen hier skizzierten Ebenen werden in vielen Fällen durch die Fachbereiche, entweder der Geistes- und Kulturwissenschaften oder der Informatik oder in Zusammenarbeit beider, verantwortet und durchgeführt. An vielen akademischen Standorten ist es aber gerade die Bibliothek, die sowohl in die Forschungsaktivitäten als auch die Lehrangebote zu den DH eingebunden ist. Dies ist vor allem in Nordamerika zu beob-

27 <http://www.i-d-e.de>

28 Die Teilnehmerzahlen stiegen kontinuierlich von 35 bei der Erstauflage 2002 bis 230 bei der zehnten Ausrichtung 2011. Im vergangenen Jahr 2012 zog das DHSI 423 Teilnehmerinnen und Teilnehmer an, ein sprunghafter Anstieg, der durch die aktuellen Anmeldezahlen für 2013 bestätigt wurde (freundliche Auskunft von Ray Siemens, E-Mail, 9.1.2013).

achten, wo traditionell die Universitätsbibliotheken weniger als ausschließliche Service-Einrichtung, sondern vielmehr als integraler Partner der akademischen Forschung und Lehre gesehen werden. Ein Beispiel hierfür ist das oben genannte Angebot der University of Virginia: das dortige „Praxis Program“ für die Lehre und das mit ihm verbundene „Scholar’s Lab“ für die Forschung sind Einrichtungen der Universitätsbibliothek (<http://www.scholarslab.org>).

Der Arbeitsmarkt

Aus Sicht der universitären Anbieter von DH-Programmen erfolgt die Ausbildung nicht im Blick auf einfache, klar vorgezeichnete Berufswege. Dies dürfte heute auch nur noch bei den wenigsten Studiengängen der Fall sein. In der Regel zielt das Lehrprofil auf einen breiten Kompetenzrahmen, der verschiedene Karrierefelder öffnen soll. Zu den Qualifikationen gehört zunächst die besondere Mischung aus einem geisteswissenschaftlichen Hintergrund und konzeptionellen sowie praktischen Anwendungskompetenzen der Informatik. Dies ist eine echte Doppel- oder Schnittstellenqualifikation, für die es sonst keine vergleichbaren Ausbildungswege gibt. Sie ermöglicht den Weg in beide Richtungen, ist aber besonders wertvoll überall da, wo sich Schnittmengen ergeben (vgl. etwa Rockwell/Sinclair 2012). Hier haben DH-Spezialisten oft eine Übersetzerfunktion zwischen den inhaltlichen Anforderungen und den technischen Lösungen. Analytische und kommunikative Fähigkeiten, verbunden mit einer tieferen kulturellen Bildung und einer Orientierung an technischer Praktikabilität machen DH-Absolventen vielfältig einsetzbar. Insgesamt lassen sich drei Bereiche unterscheiden, in denen sie besonders häufig arbeiten.

Zunächst zielt eine akademische Ausbildung immer noch auf die Befähigung, wissenschaftlich arbeiten zu können. Digitale Geisteswissenschaftler haben bislang immer noch ein Standbein in einer der geisteswissenschaftlichen Fächer und können dort eine wissenschaftliche Laufbahn einschlagen.²⁹ Zugleich entstehen immer mehr dedizierte Abteilungen und Lehrstühle für Digital Humanities, die mit den zahlreichen Forschungsprojekten in diesem Feld auch eine berufliche Identifikation als „Digital Humanist“ ermöglichen.

²⁹ Zu Berufsfeldern von Absolventen der „traditionellen“ Geisteswissenschaften siehe jüngst Kleinhagenbrock/Petersen (2011).

Der zweite Bereich lässt sich als „wissenschaftsnah“ beschreiben. Hier sind vor allem die Einrichtungen des Kulturerbes zu nennen. Bibliotheken, Archive und Museen bieten eine Reihe von Beschäftigungsmöglichkeiten in der Schnittmenge von kulturellem Erbe und digitaler Erschließung und Bereitstellung. Schließlich zeigen mindestens die Erfahrungen aus den laufenden Programmen in Köln, dass der größte Teil der Absolventen ohne größere Probleme Stellen auf dem „normalen“ Arbeitsmarkt findet. Hier sind die typischen Arbeitgeber Firmen aus den Bereichen Software, Medien oder Bildung, aber auch alle anderen Unternehmen, die im Bereich der Informationssysteme, spezifischer Softwarelösungen, des Wissensmanagements und der Beratung gerne Kandidaten einstellen, die nicht zu einseitig spezialisiert sind, sondern verschiedene Kompetenzbereiche abdecken können. Derzeit gehen jedenfalls z.B. aus den Kölner Studiengängen die allermeisten B.A.-Absolventen direkt in eine Beschäftigung. Für die international weiter verbreiteten M.A.-Studiengänge scheinen diese für die meisten Studierenden eine Etappe zur wissenschaftlichen oder wissenschaftsnahen Laufbahn, eine Ergänzung zu einer bereits bestehenden Beschäftigung oder die auf einem geisteswissenschaftlichen Studium aufsetzende Zusatzqualifikation für einen anschließenden Berufsweg zu sein.

Es ist sehr schwierig, empirisch abgesicherte Aussagen über die Berufsperspektiven der DH-Absolventen zu machen, zumal der Entwicklungsstand der verschiedenen DH-Programme höchst unterschiedlich ist. Während manche dieser Programme sich noch im Aufbau oder gar erst in der Planungsphase befinden, entwachsen andere gerade den Kinderschuhen. Ein Beispiel hierfür ist der oben genannte B.A.-Studiengang an der Universität Würzburg, in dem nach drei Jahren nun die ersten Studierenden graduiert werden. Damit bot sich die Möglichkeit, Erwartungen und Ziele vor allem von Berufsperspektiven eines neuen Studienganges aus Sicht der Studierenden in Erfahrung zu bringen und quantitativ, wohl aber nicht repräsentativ, zu unterlegen.³⁰ Die hierzu im Folgenden herangezogene Umfrage, durchgeführt im

30 Die Umfrage wurde als Projektarbeit von Susanne Strauß, Studentin in Digital Humanities sowie der Political and Social Studies, unter Betreuung von Malte Rehbein im Januar und Februar 2012 durchgeführt. 74 von 111 Studierenden (davon ⅔ weiblich und 84% jünger als 25 Jahre) des BA-Studienganges Digital Humanities haben sich per Online-Fragebogen an dieser Datenerhebung beteiligt. Eine Beschreibung dieses Projektes und eine Zusammenfassung seiner Ergebnisse ist online zugänglich: http://www.germanistik.uni-wuerzburg.de/lehrstuehle/computerphilologie/mitarbeiter/rehbein/projektarbeiten/berufsperspektiven_digital_humanities/.

dritten Jahr des Bestehens des Studienganges DH an der Universität Würzburg, erfasste die Studierenden aller Studienjahre des dreijährigen Programms und insbesondere die „erste Generation“, die zum Wintersemester 2009/10 begonnen und 2012/13 den Abschluss erworben hat oder zumindest kurz davorsteht.

Die Würzburger Studierenden der DH belegen insgesamt 18 verschiedene Fächer der Geistes- und Sozialwissenschaften, wobei Anglistik (32,4%) und Germanistik (27,0%) deutlich herausragen. Mit Ausnahme von Geografie (in Würzburg Teil der Philosophischen Fakultät), Geschichte, Kunstgeschichte und Political and Social Studies sind andere Kombinationen jedoch von höchstens ein oder zwei Studierenden belegt. Die Befragung zeigt, dass die Entscheidung, Digital Humanities (als Haupt- oder Nebenfach) zu studieren, im Wesentlichen auf der Beschreibung des Studienganges beruht, wie sie etwa online³¹ zu finden ist. Als Kriterien, die zur Wahl von Digital Humanities geführt haben, werden vorrangig gute Zukunftsaussichten, der Ausblick auf eine gut bezahlte Berufstätigkeit, die Möglichkeit, moderne Techniken kennenzulernen und „etwas mit Medien zu machen“ genannt – wobei der letzte Grund, also über das Studium in den Bereich Journalismus und andere Medien einsteigen zu können, wohl an einer zu Beginn noch nicht ausreichend kommunizierten Zielsetzung des Studienganges, vielleicht des Faches DH an sich, liegt.³² Betrachtet man die Gründe, die zur Wahl dieses Studiums führten, ist die Frage des späteren Nutzens (Berufsaussichten) vorrangig unter den zur Auswahl stehenden Antworten benannt. Andere Gründe, wie etwa, dass es sich um einen kleinen Studiengang mit guter Betreuung durch die Dozenten handelt, spielen anscheinend nur eine nachrangige Rolle bei der Entscheidungsfindung.

Die Digital Humanities erscheinen noch nicht als klar profiliertes Fach. Durch allgemeine Kenntnisse oder die Schulbildung können noch keine präzisen Vorstellungen bestehen. Wie stellen Interessenten sich dann ein Studium vor, das man nur durch eine Beschreibung, nicht aber durch Anschauung kennenlernt? Allgemein erwarteten die Würzburger Studierenden vor Aufnahme ihres Studiums ein „Arbeiten am Computer“, waren dann jedoch

31 auf den Seiten der Zentralen Studienberatung: <http://www.uni-wuerzburg.de/?id=87781>

32 Nach Auskunft der Würzburger Studienberatung wurde der Studiengang „Digital Humanities“ in seiner Anfangsphase häufig mit dem Studiengang „Medienkommunikation“ verwechselt.

vom geforderten Informatikanteil des Studiums überrascht.³³ Das ist eine auffällige Diskrepanz zwischen den Erwartungen an den Studiengang und seinen tatsächlichen Anforderungen. Diese lässt sich auch als Unterschied zwischen dem Anwenden von Computertechnologie auf der einen und dem Verstehen – im Sinne theoretischer Reflexion oder gar Weiterentwicklung – von Computertechnologie auf der anderen Seite beschreiben.³⁴ Unter den sonstigen Erwartungen wurden am häufigsten genannt: breites Themenspektrum, umfassende Ausbildung – aber mit der Möglichkeit individuell festlegbarer Schwerpunkte; vielfältiger Einblick in die Breite der Geisteswissenschaften und eine innovative Ausrichtung des Studiums. Besonders stark wird zudem Praxisnähe erwartet. Die Analyse zeigt jedoch, dass aus Sicht der Würzburger Studenten die meisten Erwartungen nur zum Teil erfüllt werden; insbesondere die angebotene fachliche Breite liegt unter den Erwartungen bzw. wird nicht wahrgenommen.

Die Befragung der Würzburger Studierendenschaft zeigt eine klare Tendenz, das Studium als Einstieg in einen konkreten Beruf zu sehen: Nur eine Minderheit von 8% stellt sich einen Berufseinstieg außerhalb ihrer beiden Studienfächer vor. Dies ist insbesondere im Bereich Digital Humanities wegen des noch fehlenden klar umrissenen Berufsfeldes bemerkenswert. Untermuert wird die Erwartungshaltung, dass das Studium auf diese Berufsfelder vorbereiten soll, von den Befragten jedenfalls durch den mehrfach angeführten Verweis auf die Online-Beschreibung des Studienganges: „Absolventen des Studiengangs können organisatorische, beratende und Managementaufgaben im Bereich der Print- und elektronischen Medien wahrnehmen und in Forschung, Journalismus, Public Relations und Werbung tätig sein.“³⁵ Als denkbare Arbeitgeber stehen die Kategorien Unternehmen und öffentliche Einrichtungen an erster Stelle, mit etwas Abstand folgen Forschung & Entwicklung sowie die Selbstständigkeit; weniger als ein Fünftel der Studierenden positioniert sich in dieser Frage noch nicht. Die Frage nach konkreten Chancen auf dem Arbeitsmarkt mit einem Abschluss in Digital Humanities

33 Diese Veranstaltung „Einführung in die Informatik für Hörer aller Fakultäten“ – angeboten durch die Fakultät für Mathematik und Informatik, im Studienverlaufsplan im 3. Semester empfohlen – hat sich als eine große Hürde für viele Studierende herausgestellt.

34 Zu dieser Frage vgl. Ramsay/Rockwell (2012).

35 <http://www.uni-wuerzburg.de/?id=87781>

werden für ein reines B.A.-Studium zurückhaltend eingeschätzt;³⁶ die Studierenden erwarten allerdings eine deutliche Verbesserung ihrer Chancen durch ein Master-Studium. Dabei geben 38% an, potenzielle Arbeitgeber bereits benennen zu können.

Die Würzburger Befragung zeigt einen „Wunsch nach einer Vielfalt an Projekten und Gruppenarbeiten, vor allem in Feldern, die auch außerhalb der Forschung liegen und in Verbindung mit potentiellen Arbeitgebern stehen“³⁷. 65% der Studierenden würden dazu ein Pflichtpraktikum während des Studiums begrüßen, allerdings macht nur eine Minderheit freiwillig ein oder mehrere Praktika im Laufe des Studiums. Diese Situation kann nur von den jeweiligen Studiengangverantwortlichen in Zusammenarbeit mit den nicht-universitären Akteuren im DH-Umfeld verändert werden. Durch interdisziplinäre Zusammenarbeit unter Einbeziehung eben nicht nur der akademischen Fachbereiche, sondern auch der Bibliotheken³⁸ und anderer kulturnaher Einrichtungen wäre dabei eine größere fachliche Breite, ein stärkerer Praxisbezug und die Eröffnung zusätzlicher Karrierechancen im „alt ac“-Umfeld (alternative academic careers) zu erreichen.³⁹

Fazit

Die Digital Humanities blicken auf eine lange Tradition zurück, können aber trotzdem als „junges“ Fach beschrieben werden, das weitgehend noch im Aufbau begriffen ist und sich in der Phase der Etablierung befindet. Da die Geistes- und Kulturwissenschaften an sich aber immer mehr von digitalen Aspekten der Informationsbereitstellung, -verarbeitung, -analyse und -auf-

36 Dies steht in direktem Gegensatz zu den o.a. Kölner Erfahrungen.

37 Siehe den Abschlussbericht (wie Anm. 30), dort S. 14.

38 Ein Beispiel hierfür stellt die Bibliothek der University of Illinois at Urbana-Champaign dar, die DH auf der „Graduate“-Ebene in den Bereich „Service“ der Bibliothek einbettet: „Faculty and graduate students embarking on a digital humanities project can use the resources in the Scholarly Commons to make connections, explore tools and obtain data [...] such as: Building an image archive and annotating the images with metadata; Transcribing manuscripts and marking them up in TEI for text encoding and text mining. Creating a map with linked data and annotations“ (http://www.library.illinois.edu/sc/services/digital_humanities.html).

39 Siehe hierzu Robert Vegas Beitrag zu „Alt-ac and Digital Humanities Jobs“ (Vegas o.J.) sowie <http://mediacommons.futureofthebook.org/alt-ac/>.

bereitung sowie Kommunikation und Ergebnispräsentation geprägt sind, steigt auch der Bedarf an ausgebildeten Wissenschaftlern, die in der Lage sind, die methodische, modellierende, formalisierende und technische Seite der Geisteswissenschaften abzudecken, weiter. Dieser Bedarf, der auch von Seiten der Politik und Forschungsförderung klar erkannt und damit unterstützt wird, führt nicht nur dazu, dass es immer mehr Akteure im weitesten Bereich der DH gibt, sondern auch zu einer zunehmenden Bedeutung einer zielgerichteten Ausbildung. Die Grundlage dafür sind die Stabilisierung und der Ausbau spezieller – meistens fakultätsweit ausgerichteter – Abteilungen und Zentren mit entsprechenden Professuren und Mitarbeiterstellen. Von entscheidender Bedeutung ist dann aber auch die Sichtbarkeit und Erkennbarkeit des Feldes und Faches „Digital Humanities“ über den engeren Kreis der Spezialisten hinaus. DH werden erst dann als nachhaltig etabliert gelten können, wenn bei den wissenschaftspolitischen Entscheidungsträgern, an den Universitäten und in der Wirtschaft mit dem Begriff „Digital Humanities“ ein einigermaßen klares Profil verbunden werden kann. Dazu tragen die Vielzahl der einzelnen aktuellen Projekte, Workshops und Tagungen ebenso bei wie die großen Verbundprojekte (z.B. DARIAH), die für den Aufbau einer gemeinsamen Infrastruktur wirken. Dazu muss aber auch klar sein, was jemand, der einen Abschluss in „Digital Humanities“ vorweist, eigentlich gelernt hat, was seine Kompetenzen und Qualifikationen sind und was man von ihm erwarten kann. Dies geschieht nur im Wechselspiel und durch Kooperation aller Akteure im Feld und in einer Verständigung – im besten Fall auch Abstimmung – über vergleichbare Lehrinhalte und Curricula.

Literaturverzeichnis

- Alexander, Bryan; Davis, Rebecca Frost (2012): Should Liberal Arts Campuses Do Digital Humanities? Process and Products in the Small College World. In: Gold, Matthew (Hrsg.): *Debates in the Digital Humanities*. Minneapolis, MN: University of Minnesota Press, S. 368–389.
- Gold, Matthew (2012): *Debates in the Digital Humanities*. Minneapolis, MN: University of Minnesota Press. Online: <http://dhdebates.org/>.
- Haber, Peter (2011): *Digital Past. Geschichtswissenschaft im digitalen Zeitalter*. München: Oldenbourg.
- Heinevetter, Nele; Sanchez, Nadine (2008): *Was mit Medien ... Theorie in 15 Sachgeschichten*. Paderborn: Wilhelm Fink.

- Hirsch, Brett D. (Hrsg.) (2012): *Digital Humanities Pedagogy: Practices, Principles and Politics*. Cambridge: Open Book Publishers.
- Kleinlagenbrock, Frank; Petersen, Stefan (Hrsg.) (2011): *Geschichte studiert – und dann? Berufsfelder für Historikerinnen und Historiker sowie Studierende anderer Geisteswissenschaften – Ein Leitfaden*. Würzburg: Universität. <http://opus.bibliothek.uni-wuerzburg.de/volltexte/2012/6670/>.
- O'Donnell, Daniel (2012): Global Outlook: Digital Humanities, Blogbeitrag vom 12.11.2012. <http://dpod.kakelbont.ca/category/projects-and-societies/godh/>.
- Ramsay, Stephen; Rockwell, Geoffrey (2012): Developing Things: Notes toward an Epistemology of Building in the Digital Humanities. In: Gold, Matthew (Hrsg.): *Debates in the Digital Humanities*. Minneapolis, MN: University of Minnesota Press, S. 75–84.
- Reid, Alexander (2012): Graduate Education and the Ethics of the Digital Humanities. In: Gold, Matthew (Hrsg.): *Debates in the Digital Humanities*. Minneapolis, MN: University of Minnesota Press, S. 357 ff.
- Rockwell, Geoffrey; Sinclair, Stéfan (2012): Acculturation in the Digital Humanities Community. In: Hirsch, Brett (Hrsg.): *Digital Humanities Pedagogy: Practices, Principles and Politics*. Cambridge: Open Book Publishers, S. 177–212.
- Schreibman, Susan; Siemens, Ray; Unsworth, John (Hrsg.) (2004): *A Companion to Digital Humanities* (Blackwell Companions to Literature and Culture; 26). Oxford: Blackwell. Online: <http://www.digitalhumanities.org/companion>.
- Vega, Robert (o.J.): Alt-ac and Digital Humanities Jobs, In: *Liberal Arts Career Services*. <http://www.utexas.edu/cola/orgs/lacs/Students/Graduate-Students/Career-Guides/Alt-ac.php>.

Hochschule im digitalen Zeitalter

Die Stärkung von Informationskompetenz als Ziel einer Empfehlung der Hochschulrektorenkonferenz

von Ulrich Meyer-Doeringhaus

Die zunehmende Bedeutung der digitalen Information verändert die Hochschule auf allen Ebenen: Die Lehre ist inzwischen ohne die elektronische Dimension undenkbar, die Kommunikation in der Forschung dringt durch Virtuelle Forschungsumgebungen, Data Sharing und elektronisches Publizieren in ganz neue Dimensionen vor und auch für die Hochschulverwaltung eröffnet ein integriertes Informationsmanagement neue Möglichkeiten. Die digitale Information ist eine Herausforderung, die die ganze Hochschule betrifft und insbesondere die Hochschulleitungen auf den Plan ruft. Aus diesem Grund hat sich die Arbeitsgruppe „Zukunft der digitalen Information in Lehre und Forschung“¹ der Hochschulrektorenkonferenz (HRK) unter Leitung von Vizepräsident Prof. Dr. Joachim Metzner der Fragen angenommen, die mit der Digitalisierung der Information im Einzelnen verbunden sind. Sie hat dabei die Herausbildung und Stärkung von Informationskompetenz auf allen Ebenen der Hochschulen als die entscheidende Herausforderung erkannt. Wie Informationskompetenz an den Hochschulen konkret gestärkt werden kann, hat die AG in der Empfehlung „Hochschule im digitalen Zeitalter: Informationskompetenz neu begreifen – Prozesse anders steuern“ beschrieben, die von der Mitgliederversammlung der HRK am 20. November 2012 in Göttingen verabschiedet worden ist.

¹ Mitglieder der Arbeitsgruppe sind: Frau Prof. Dr. Petra Gehring (TU Darmstadt), Frau Prof. Dr. Ulrike Gutheil (TU Berlin), Herr Prof. Dr. Wilfried Juling (Karlsruhe Institute of Technology), Frau Dr. Antje Kellersohn (Universität Freiburg), Herr Prof. Dr. Norbert Lossau (Universität Göttingen), Herr Dr. Frank Nolden (Universität Leipzig), Herr Prof. Dr. Joachim Schachtner (Universität Marburg), Herr Prof. Dr. Peter Schirnbacher (Humboldt-Universität zu Berlin), Herr Prof. Dr. Rolf Schwartmann (FH Köln), Herr Prof. Dr. Uwe Schwiegelshohn (TU Dortmund), Frau Dr. Beate Tröger (Universität Münster) und Herr Prof. Martin Ullrich (Hochschule für Musik Nürnberg).

Die Empfehlung ist nicht die erste Äußerung der HRK zum angesprochenen Thema. Schon 2006 hatte sie mit einem „Leitfaden für Hochschulstrategien zur Informations- und Kommunikationskultur“ den Hochschulleitungen und den leitenden Verantwortlichen der IT-Gesamtstruktur Hilfestellungen für die Formulierung einer hochschulspezifischen Strategie für die Informations- und Kommunikationskultur einer Hochschule an die Hand gegeben und darin die Umsetzung entsprechender Projekte beispielhaft beschrieben. Daran hat die Handreichung „Herausforderungen Web 2.0“ aus dem Jahr 2010 anknüpfen können, die die bereits vorhandenen Potenziale der Hochschulen und ihre Chancen mit Blick auf das Web 2.0 beschrieben hat. Sofern dieser Text auch auf den Erwerb von Informationskompetenz eingeht, ist er als wichtige Voraussetzung der jüngst verabschiedeten HRK-Empfehlung anzusehen.

Das Thema fügt sich zugleich sehr gut in einen Sammelband zum zehnjährigen Bestehen der Abteilung Forschung & Entwicklung der Universitäts- und Landesbibliothek Göttingen ein. Bekanntlich gehen von dieser Abteilung wichtige Impulse für die Entwicklung der wissenschaftlichen Infrastrukturen an Hochschulen aus. In vielerlei Hinsicht haben die Göttinger eine treibende Funktion für die Entwicklung im Hochschulsystem. Was speziell das Thema Informationskompetenz angeht, sei vor allem auf die verdienstvollen Aktivitäten mit Blick auf das Thema Langzeitarchivierung hingewiesen. Die Abteilung ist etwa an „nestor“, dem deutschen Kompetenznetzwerk zur digitalen Langzeitarchivierung, maßgeblich beteiligt. Im Rahmen des Teilprojekts „nestor Qualifizierung“ werden wichtige Anregungen zur Lösung der Frage erarbeitet, wie Informationskompetenz mit Blick auf die Langzeitarchivierung gestärkt werden kann. Außerdem muss auf das Engagement der Abteilung mit Blick auf die Beschreibung der Anforderungen hingewiesen werden, die sich für neue Dienstleistungsprofile, etwa im Bereich des Forschungsdatenmanagements, ergeben (vgl. Engelhardt/Strathmann/McCadden 2012; Oßwald/Strathmann 2012).

Im Folgenden wird zunächst beschrieben, welche Voraussetzungen die Empfehlung der HRK im wissenschaftspolitischen Umfeld hatte und wie die HRK dazu gekommen ist, sich des Themas anzunehmen. Sodann seien die wichtigsten Aussagen der Empfehlung skizziert.

Voraussetzungen und Vorgeschichte der HRK-Empfehlung

Kultur des Internets

Die Empfehlung blickt zunächst bewusst über den Bereich der Hochschule hinaus, indem sie sich auf gesamtgesellschaftliche Entwicklungen bezieht:

„Kommunikation erlebt derzeit rasante Veränderungen, die vor allem durch Innovationen im Bereich der digitalen Information angestoßen werden. Die wachsende Bedeutung der sog. social media zeigt, dass es dabei nicht nur um technologische Veränderungen geht. Es entstehen neue soziale Strukturen und Gewohnheiten, die mit neuen Sprach-, Werte- und Orientierungsmustern einhergehen. Auch geht es um einen neuen, im höheren Maße sinnlichen und partizipativen Umgang mit Information.“ (HRK 2012: 4)

Bekanntlich machen diese Entwicklungen vor den Hochschulen nicht halt und haben Auswirkungen auf alle Bereiche einer Hochschule. So sehr die Vorteile der Digitalisierung von Informationen auf der Hand liegen, so sehr tun sich für die Hochschulen Herausforderungen großen Umfangs auf, die längst noch nicht bewältigt sind:

„Die Studierenden erwarten etwa, dass die Hochschulen mit den neuen Entwicklungen Schritt halten und ihnen eine integrierte Informationsumgebung zur Verfügung stellen, die den außerhalb der Hochschule eingeübten Gewohnheiten Rechnung trägt. Netzgestützte Lehrangebote gewinnen für die Studierenden an Relevanz. Zugleich tragen die Studierenden, die an die Kultur des Internets gewöhnt sind, neue Fragen an das Wissenschaftssystem heran: Wie etwa ist die für das Wissenschaftssystem maßgebliche Wertschätzung der individuellen Forschungsleistung als eines geistigen Eigentums mit den Konventionen vereinbar, die in sozialen Netzwerken praktiziert werden? Auch die Forschung verändert sich grundsätzlich: Wissen wird in der Interaktion ständig neu produziert und muss im Modell des information life cycle neu begriffen werden. Die Digitalisierung von Forschungsdaten schreitet voran. Datenmengen wachsen exponentiell, neue Lösungen des Managements, der Speicherung und Archivierung sind gefragt. Mit neuen Möglichkeiten geht einher, dass einige der Prozesse, die früher das ‚gebildete Individuum‘ gesteuert hat, heute bereits automatisiert sind. Das wiederum wirft neue Fragen mit Blick auf Ethik und Verantwortung in der Forschung auf.“ (ebd.: 4 f.)

Der allgemeine und übergreifende Charakter dieser Herausforderungen weist vor allem den Hochschulleitungen die Aufgabe zu, Lösungen herbeizuführen. Nur aus der Leitungsperspektive können diese erkannt, entworfen und – im Rahmen eines Governance-Modells – realisiert werden. Deshalb wendet sich die Empfehlung in erster Linie an die Hochschulleitungen, in zweiter Linie

an die unterschiedlichen, vom Thema des Papiers betroffenen Teilbereiche der Hochschule.

Gesamtkonzept der „Kommission für Informationsinfrastruktur“ (KII)

Eine weitere wichtige Voraussetzung des HRK-Papiers stellen die Empfehlungen der „Kommission für Informationsinfrastruktur“ (KII) dar, die vom Ausschuss der Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder (GWK) im Mai 2011 zustimmend zur Kenntnis genommen worden sind. Das Kapitel zur Informationskompetenz, die im Verhältnis zu allen anderen Handlungsfeldern (z. B. „Lizenzierung“, „Hosting/Langzeitarchivierung“, „Nichttextuelle Materialien“ etc.) ein zentrales Querschnittsthema darstellt, ist von der HRK federführend verantwortet worden. Die HRK ist dabei im Text als künftige bundesweite Koordinatorin für das Thema „Informationskompetenz“ benannt worden (vgl. KII 2011: 65).

Empfehlungen des Wissenschaftsrates

Das HRK-Papier fußt zudem auf den unterschiedlichen Papieren des Wissenschaftsrats zum Themenkreis, insbesondere auf den „Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020“ vom Juli 2012. In diesem Papier nimmt der Wissenschaftsrat zustimmend Stellung zum KII-Gesamtkonzept und empfiehlt u. a. den „Aufbau von spezifischer Medien- und Informationskompetenz (Medienbildung)“ (Wissenschaftsrat 2012: 41) an den Hochschulen. Schon Anfang 2011 hatte der Wissenschaftsrat in seinen „Übergreifenden Empfehlungen zu Informationsinfrastrukturen“ dafür plädiert, die Entwicklung der wissenschaftlichen Informationsinfrastrukturen oben auf die wissenschaftspolitische Agenda zu setzen und entsprechende förderpolitische Instrumente zu realisieren.

Die Inhalte der Empfehlung

Begriffliche Vorklärungen

Der Begriff „Informationskompetenz“ wird mit Blick auf die Hochschule üblicherweise allein auf eine Reihe akademischer Qualifikationen im Bereich des Lehrens, Lernens und Forschens bezogen. Die Empfehlung der HRK erweitert das Gegenstandsfeld des Begriffs dagegen erheblich:

„Auch die Leitung einer Hochschule und die Dienstleistungen, mit denen Forschung und Lehre unterstützt werden, werden als Gegenstandsbereiche von Informationskompetenz betrachtet. Auf diese Weise soll deutlich gemacht werden, dass die akademische Kerntätigkeit einer Hochschule, also das Lehren, Lernen und Forschen, in enger Verzahnung mit der Hochschule als Organisation zu sehen ist und dass alle Anstrengungen zur Stärkung von Informationskompetenz in der Hochschule sich sowohl auf den akademischen Kernbereich als auch auf die Organisation richten müssen.“ (HRK 2012: 5 f.)

Auf dieser Grundlage kommt die Empfehlung zu einer Definition von „Informationskompetenz“, die auf einer Begriffsbestimmung von Prof. Dr. Christian Wolff (Universität Regensburg) beruht. Nach diesem ist Informationskompetenz

„die Gesamtheit aller Fähigkeiten und Fertigkeiten, die erforderlich sind, um situationsrelevante Informationsbedarfe festzustellen, Information zu beschaffen, weiterzuverarbeiten, zu bewerten, zu präsentieren und Nutzungsbedingungen von Information einzuordnen. Dabei sind neue Darstellungsformate, insbesondere im Bereich der Informationsvisualisierung, eingeschlossen.“ (ebd.: 6)

Dem fügt die Empfehlung noch einen weiteren Aspekt hinzu, der angesichts wachsender Datenmengen und rechtlicher Grauzonen zunehmend an Aktualität gewinnt, nämlich die „Informationsverantwortung“. Sie beinhaltet „ein Bewusstsein für Chancen, Gefahren und Risiken, die mit dem Umgang mit Informationen verbunden sind“ (ebd.). Außerdem grenzt sich die Empfehlung vom herkömmlichen, stark bibliothekswissenschaftlich geprägten Verständnis von Informationskompetenz als einer reinen Recherchekompetenz ab, indem sie betont, dass Informationskompetenz „heute in besonderer Weise auf den Umgang mit den neuen Entwicklungen der wissenschaftlichen Informationsinfrastrukturen (z. B. mit Virtuellen Forschungsumgebungen und Forschungsdaten) zu beziehen ist“ (ebd.). Auf dieser Grundlage wird sodann der allgemeine Begriff mit konkreten Inhalten gefüllt, indem, wiederum an Wolff anknüpfend, eine Reihe von Teilfertigkeiten genannt werden, nämlich

- eine technische Kompetenz, d. h. ein für die Anwendung verschiedener Informations- und Kommunikationsmedien erforderliches technisches Wissen (als Weiterführung der sog. computer literacy),
- eine kommunikative Kompetenz, d. h. ein Wissen um die Verfügbarkeit und Funktion der digitalen Kommunikationsmedien,
- eine soziale und organisationsbezogene Kompetenz sowie
- eine disziplinspezifische Kompetenz, d. h. ein Wissen um die Besonderheiten unterschiedlicher Wissenschaftskulturen.“ (ebd.)

Im Folgenden kommt die Empfehlung zu einer Klarstellung, die einem häufig geäußerten Missverständnis begegnen soll. In Zeiten der zunehmenden Relevanz der digitalen Information wird „Informationskompetenz“ oftmals gleichgesetzt mit der Kompetenz, elektronische Informationen und Medien adäquat nutzen zu können. Dem hält die Empfehlung einen Einwand entgegen, der auf jüngeren Erfahrungen im E-Learning-Bereich beruht: Nicht-digitale Informationsträger behalten vor allem in der Lehre ihre ungebrochene Relevanz. Deshalb kann „Informationskompetenz“ mitnichten allein auf den adäquaten Umgang mit digitaler Information bezogen werden: „Vielmehr kann gerade ein ausgewogener Umgang mit digitalen und nicht-digitalen Informationsträgern (etwa in der Lehre) besonderer Ausdruck von Informationskompetenz sein“ (ebd.).

Was das inhaltliche Verhältnis zwischen dem Begriff „Informationskompetenz“, der dem englischen Wort *information literacy* entlehnt ist, und dem im Deutschen weit üblicheren Terminus der „Medienkompetenz“ angeht, zieht die Empfehlung eine scharfe Grenze:

„Gegenüber dem Begriff der ‚Medienkompetenz‘, der stärker auf die adäquate Nutzung neuer Informations- und Kommunikationstechnik abhebt, wird in diesem Papier bewusst dem übergeordneten Begriff der ‚Informationskompetenz‘ der Vorzug gegeben. Dieser Terminus, der die bestmögliche Organisation von Informationen als individuelle menschliche Leistung in den Mittelpunkt stellt, geht über die optimale Handhabung bestimmter Medien hinaus und lenkt die Aufmerksamkeit auch auf Infrastrukturen. Er hat außerdem soziale, kulturelle und ethische Implikationen. Gleichwohl schließt der Begriff ‚Informationskompetenz‘ den Begriff der ‚Medienkompetenz‘ mit ein.“ (ebd.: 7)

Studium und Lehre

Mit Blick auf Studium und Lehre unterscheidet die Empfehlung einerseits zwischen der Informationskompetenz, die als Ziel der Lehre den Studierenden in Seminaren, Vorlesungen etc. zu vermitteln ist, und andererseits der Informationskompetenz, die den Lehrenden idealerweise zur Verfügung stehen sollte. In Bezug auf den zunächst genannten Begriffsinhalt wird hervorgehoben, dass die Studierenden bereits ein hohes Maß an selbst erlernter Informationskompetenz aus ihrer Erfahrung mit dem Netz in die Hochschule mitbringen. Diese Voraussetzung muss die Lehre dringend berücksichtigen, will sie ihre Ziele nicht verfehlen. Wenn die Lehre nicht redundant sein will, d. h. den Studierenden nicht bereits Vertrautes vermittelt, muss sie besonders auf die fachspezifische Informationskompetenz abstellen: Fakultäten und Fachbereiche, Bibliotheken, Rechenzentren und Medienzentren sollten unter

dieser Prämisse dafür Sorge tragen, dass Studierende lernen, „Informationen technisch aufzubereiten, sie in ihre persönliche Informationsumgebung zu übernehmen, sie in eine ansprechende Informationsoberfläche zu integrieren, sie didaktisch ausgereift zu präsentieren und sie zu archivieren“ (ebd.: 9).

Die Empfehlung umreißt im Folgenden drei große Herausforderungen bei der Vermittlung von Informationskompetenz an die Studierenden:

„1. Die Lehrangebote müssen die Kompetenzen berücksichtigen und ergänzen, die die Studierenden bereits in anderen Kontexten – vor allem durch Selbstlernen im Netz – erworben haben. Weit mehr als früher lernen junge Leute bereits an der Schule die Möglichkeiten des Internets kennen und beschreiten während ihres Studiums Wege des informellen Lernens. Lehrende sollten deshalb darauf achten, dass sie keinen redundanten oder gar unterfordernden Ansatz wählen. Vielmehr sollten bei der Vermittlung von Informationskompetenz solche Inhalte im Vordergrund stehen, die im Netz nicht erlernbar sind.

2. Oft findet keine genügende Abstimmung zwischen den Angeboten der verschiedenen Akteure einer Hochschule statt. Deshalb ist es notwendig, die Angebote stärker als bisher aufeinander abzustimmen. Ein größeres Zusammenwirken der Akteure beinhaltet besondere Potenziale zum Vorteil der Studierenden, weil die Akteure unterschiedliche, sich ergänzende Perspektiven und Stärken einbringen.

3. Eine zentrale Herausforderung besteht darin, die Lehrangebote zur Vermittlung von Informationskompetenz stärker als bisher curricular zu verankern. Bislang werden jene zum deutlich überwiegenden Teil außerhalb des regulären Curriculums angeboten. Dagegen sollten die Angebote stärker als zuvor verpflichtend gemacht werden, damit eine flächendeckende Stärkung der Informationskompetenz bei den Studierenden ermöglicht wird. Es sollte vor allem an die Einführung neuer Studiengangmodule oder die Integration neuer Lehrangebote in bereits vorhandene Module gedacht werden.“ (ebd.: 10)

Was den letzten Punkt angeht, sollten bei der Konzeption solcher Module bzw. Modulteile bestimmte Gesichtspunkte besonders beachtet werden. Dazu gehört etwa, im grundständigen Studium dem allgemeinen bzw. generischen Anteil, im weiterführenden Studium dagegen dem fachspezifischen Anteil mehr Raum zu geben. Im Masterstudium sollte insbesondere die forschungsbezogene Informationskompetenz (d.h. z.B. die Kenntnis der Grundlagen des Informationsmanagements und der neuen Entwicklungen des digitalen Forschungsprozesses) in Grundlagen vermittelt werden.

Mit Blick auf die Informationskompetenz der Lehrenden unterstreicht die Empfehlung, dass jene selbst die Kompetenzen erwerben müssen, die sie vermitteln, und zugleich um die Informationskompetenz, die Studierende vor und außerhalb ihres Studiums erworben haben, wissen müssen. Dabei umfas-

sen die notwendigen Kompetenzen auch die neuen Gegenstände des E-Learning:

„Die digitale Dimension der Lehre schließt dabei ein, auch die Möglichkeiten des Web 2.0 und die damit verbundenen interaktiven Lehr- und Lernszenarien nutzen zu können. Eine informationskompetente Lehre reicht außerdem von der richtigen Entscheidung darüber, unter welchen Bedingungen die Anwendung von Literaturverwaltungsprogrammen (auch in den Geisteswissenschaften) sinnvoll ist, über die optimale Einrichtung und Nutzung eines elektronischen Semesterapparates bis hin zur Bereitstellung von Veranstaltungsmitschnitten mit ergänzenden Informationen in Repositorien.“ (ebd.: 11)

Vor diesem Hintergrund wird gefordert, dass die Lehrenden erforderlichenfalls entsprechende Fortbildungs- und Trainingsangebote wahrnehmen sollten. Dies verlange entsprechende Anstrengungen auch der Hochschulleitungen, welche „ihrerseits eine ausreichende Personalausstattung und die hohe Qualität des Personals in den qualifizierenden Einrichtungen sicherstellen“ (ebd.: 12) sollten.

Forschung

Bekanntlich führt die digitale Information auch in der Forschung zu tiefgreifenden Veränderungen des Arbeitsprozesses, werden doch die Forschungsinformation und die Kommunikation in der Forschung auf ganz neue Grundlagen gestellt. Es sei nur auf die fortschreitende Digitalisierung von Texten und Objekten, das elektronische Publizieren und Open Access, Virtuelle Forschungsumgebungen, Forschungsdaten und Langzeitarchivierung verwiesen. Die HRK-Empfehlung sieht darin eine irreversible Entwicklung und fordert deshalb einen nachhaltig wirksamen Ausbau der Informationskompetenz:

„Auch wenn die neuen Entwicklungen je nach Fächerkultur in unterschiedlichen Geschwindigkeiten adaptiert werden, so ist von einem unumkehrbaren Trend auszugehen: Die Anwendung der neuen Möglichkeiten dürfte in naher Zukunft in den meisten Fächerkulturen selbstverständlich sein. Die Stärkung der Informationskompetenz sollte deshalb zu einem zentralen Baustein der Förderung des wissenschaftlichen Nachwuchses werden. Es ist notwendig, die Forscherinnen und Forscher bei der Entwicklung der forschungsbezogenen Informationskompetenz mehr als bisher zu unterstützen und ihnen entsprechende Qualifizierungsangebote zur Verfügung zu stellen. Bestehende Angebote sollten stärker als bisher in den Curricula der Graduierten- und Postgraduiertenausbildung verankert werden.“ (ebd.)

Von zentraler Bedeutung erscheint aber nicht nur die individuelle Förderung der Wissenschaftlerinnen und Wissenschaftler. Vielmehr sollte Ziel der Be-

strebungen sein, bestehende, aus der Initiative der Forschung selbst hervorgegangene Netzwerke zu unterstützen. Dabei sollte berücksichtigt werden, dass auch nicht-wissenschaftliche Partner aus angrenzenden gesellschaftlichen Feldern eine wichtige Rolle in solchen Netzwerken spielen können. In der Linie eines entsprechenden Bottom-up-Ansatzes liegt auch der folgende Ratschlag der Empfehlung. Es sei

„sinnvoll, für jedes der oben genannten Themenfelder (z. B. Elektronisches Publizieren und Open Access, Virtuelle Forschungsumgebungen) je eigene Kompetenznetzwerke zu fördern bzw. neu zu initiieren. Deren Koordination müsste bei einer Einrichtung liegen, die mit Blick auf das jeweilige Themenfeld im bundesweiten Vergleich eine herausragende Expertise aufweist.“ (ebd.: 13)

Einen weiteren Lösungsvorschlag spricht die Empfehlung sodann mit Blick auf eine spezielle, aber nicht minder drängende Herausforderung aus: Forscherinnen und Forscher sollten, um die neuen Möglichkeiten der digitalen Information adäquat nutzen zu können, bestimmte Basiskompetenzen entwickeln, d.h. „grundlegende Kenntnisse des Datenmanagements, eine allgemeine juristische Kompetenz mit Blick auf das Urheberrecht und eine besondere Kompetenz mit Blick auf die adäquate Präsentation von Forschungsergebnissen“ (ebd.). Zweifellos sollten sich aber Forscherinnen und Forscher nicht zu IT-Spezialisten entwickeln – dazu fehlen ihnen allein schon die zeitlichen Ressourcen. Aus diesem Grund sollten sie sich, so die Empfehlung, bei der Nutzung neuer digitaler Instrumente vom Personal der Bibliotheken und Rechenzentren mehr als bisher unterstützen lassen. Das allerdings zwingt die hochschulinternen Dienstleister zu bedeutenden Veränderungen: Die von ihnen angebotenen Services für die Forschung müssen mehr als bisher auf die spezifischen Anforderungen der Forschung zugeschnitten werden.

Einen besonders akuten Bedarf sieht die Empfehlung deshalb mit Blick auf das Management von Forschungsdaten. Sie stellt zwei Kompetenzerweiterungen vor, die sich derzeit in der Diskussion befinden:

„die des ‚Data Librarian‘ (eine hochqualifizierte, auch fachwissenschaftlich ausgebildete Informationsfachkraft, die im engen Austausch mit den Wissenschaftlerinnen und Wissenschaftlern Daten organisiert und dadurch den Forschungsprozess unmittelbar unterstützt) und die des ‚Data Curator‘ (eine Person, die das technische Datenmanagement betreut).“ (ebd.)

Mit Blick auf die bereits ältere Debatte über das ideale Verhältnis zwischen Hochschulbibliotheken und -rechenzentren erscheint dabei bemerkenswert, dass die neuen beruflichen Profile die hergebrachte Trennung der beiden Dienstleistungsbereiche obsolet erscheinen lässt, da die neuen Spezialistin-

nen und Spezialisten an der Schnittstelle nicht nur zwischen Wissenschaft und Infrastruktureinrichtungen, sondern auch zwischen den Infrastruktureinrichtungen selbst agieren.

Um ein Angebot zur Entwicklung der neuen Profile aufzubauen, das einen quantitativ nennenswerten Umfang hat, empfiehlt die HRK ihren Mitglieds-hochschulen,

„den genannten Ausbau der Kompetenzen durch die Erweiterung bestehender grundständiger Studiengänge und die Einführung neuer Angebote im Bereich der wissenschaftlichen Weiterbildung (z. B. Zertifikatskurse) im Themenfeld des Datenmanagements engagiert zu begleiten. Zugleich sollten entsprechende Maßnahmen der Personalentwicklung an den Bibliotheken und Rechenzentren aktiv unterstützt werden.“ (ebd.: 16 f.)

Der Aufbau von Strukturen und Fragen der Finanzierung

Welche strukturellen Maßnahmen sind erforderlich, um die zuvor genannten Empfehlungen erfolgreich und nachhaltig umzusetzen? Die HRK-Empfehlung setzt weniger auf den Ausbau neuer Netzwerke und Suprastrukturen als vielmehr auf die „Unterstützung bestehender oder Einrichtung neuer Netzwerke, die sich zum Teil speziellen Zielen widmen und sich regional zusammenschließen“ (ebd.: 17).

Darüber hinaus werden folgende Lösungsansätze vorgestellt, an denen sich die HRK zum Teil unmittelbar beteiligt sieht:

- Bereits im Rahmen des sog. KII-Prozesses hatte sich die HRK bereit erklärt, sich als bundesweite Koordinatorin für das Handlungsfeld „Informationskompetenz“ in der Wissenschaft zur Verfügung zu stellen (vgl. KII 2011: 66). Entsprechend wird die HRK ihre Bemühungen fortsetzen, ein Netzwerk aufzubauen, das Vorschläge zur Frage unterbreitet, wie Informationskompetenz besser in die Studiengänge integriert werden kann. Außerdem sollte das Netzwerk den Stand der Verankerung von Informationskompetenz an den Hochschulen fortlaufend beobachten.
- Die HRK erklärt sich in der Empfehlung außerdem dazu bereit, eine Arbeitsgruppe zu gründen, die die Prozesse an der Schnittstelle zwischen den Fakultäten bzw. Fachbereichen einerseits und den Bibliotheken und Rechenzentren andererseits mit Blick auf ausgewählte Fächer in regelmäßigen Zeitabständen begleitet und Vorschläge zur Verbesserung der Zusammenarbeit formuliert.
- Mit Blick auf die entscheidende Frage der Finanzierung der empfohlenen Maßnahmen ruft die Empfehlung die Politik in Bund und Ländern dazu

auf, „ihren Anteil am Aufbau solcher Strukturen zu leisten und entsprechende Maßnahmen finanziell zu unterstützen“ (HRK 2012: 18). Es liegt nahe, dass die dringend notwendige Stärkung von Informationskompetenz auf allen Ebenen der Hochschule nur dann wirkungsvoll in Angriff genommen werden kann, wenn es zu einem entsprechenden Bekenntnis der Politik kommt, das sich in der Finanzierung vielversprechender Projekte niederschlägt.

Literaturverzeichnis

- Engelhardt, Claudia; Stefan Strathmann; Katie McCadden (2012): Report and analysis of the survey of Training Needs. <http://www.digcur-education.org/eng/content/download/3322/45927/file/Report%20and%20analysis%20of%20the%20survey%20of%20Training%20Needs.pdf>.
- KII (Hrsg.) (2011): Gesamtkonzept für die Informationsinfrastruktur in Deutschland, hrsg. v. Kommission Zukunft der Informationsinfrastruktur. http://www.allianz-initiative.de/fileadmin/user_upload/KII_Gesamtkonzept.pdf.
- KRK (Hrsg.) (2012): Hochschule im digitalen Zeitalter: Informationskompetenz neu begreifen – Prozesse anders steuern, Entschließung der 13. Mitgliederversammlung der HRK am 20. November 2012 in Göttingen. http://www.hrk.de/uploads/tx_szconvention/Entschl._Informationskompetenz_final_20_11.pdf, im Printformat kostenlos zu beziehen unter: <http://www.hrk.de/publikationen/gesamtliste-hrk-publikationen/>.
- Oßwald, Achim; Stefan Strathmann (2012): The Role of Libraries in Curation and Preservation of Research Data in Germany: Findings of a survey. Vortrag auf der 78th IFLA General Conference and Assembly, Helsinki. <http://conference.ifla.org/past/2012/116-osswald-en.pdf>.
- Wissenschaftsrat (Hrsg.) (2012): Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020. <http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf>.

Technologien und Dienste

Forschungsinfrastrukturen in den Geisteswissenschaften

DARIAH-EU als europäische Forschungsinfrastruktur

von Tobias Blanke und Christiane Fritze

Einleitung: DARIAH als europäische Forschungsinfrastruktur

Forschungsinfrastrukturen für die Geisteswissenschaften sind noch jung. Während in den Naturwissenschaften schon direkte Beziehungen zwischen dem Wachsen der Disziplinen und dem Wachsen der Forschungsinfrastrukturen gezogen werden können, sind wirksame Infrastrukturen in den Geistes- und Kulturwissenschaften noch im Aufbau begriffen und selbst Teil der Forschung. Traditionell haben Bibliotheken hier eine große Rolle gespielt. Sie sind historisch der Ort der geisteswissenschaftlichen Infrastrukturen gewesen. Mit dem digitalen Wandel müssen auch diese Forschungsinfrastrukturen neu entwickelt werden. Dieser Beitrag wird von einem solchen Versuch berichten, geisteswissenschaftliche Infrastrukturen (neu) zu erfinden, zumindest für den noch recht neuen Teil der Kunst- und Geisteswissenschaften, die sich mit digitalen Methoden und Ressourcen auseinandersetzt.

DARIAH-EU ist die Digital Research Infrastructure for the Arts and Humanities. In DARIAH-EU kommen viele europäische Staaten zusammen und unterschiedliche Traditionen des akademischen Zusammenarbeitens treffen aufeinander. DARIAH-EU ist für Geistes- und KulturwissenschaftlerInnen ein sehr großes Projekt mit einem finanziellen Rahmen von vier Millionen Euro jährlich. Verglichen mit Forschungsinfrastrukturen in den Naturwissenschaften, wie z.B. dem Large Hadron Collider¹ am CERN oder Riesenteleskopanlagen, ist es eher klein. DARIAH-EU ist außerdem eine virtuelle Infrastruktur, die sich vor allem der digitalen Integration von Forschung, Forschungsgegenständen und Forschungsobjekten widmet.

Wir mussten bei DARIAH-EU ganz von vorne anfangen, wobei wir auf die Erfahrung in den verschiedenen Mitgliedsländern von DARIAH-EU zu-

¹ <http://lhc.web.cern.ch/lhc/>

rückgreifen konnten. Während DARIAH-EU aufgesetzt worden ist, haben wir miterlebt, wie einige europäische Länder ihre Anstrengungen fast ausschließlich auf digitale Methoden in den Geisteswissenschaften konzentrieren, während andere sich zunehmend aus dem Gebiet zurückziehen. Die digitalen Geistes- und Kulturwissenschaften sind eine der großen Hoffnungen für die Zukunft der Forschung, aber eben auch oft mit bisher nicht gekannten Kosten verbunden. Diese sind für viele in den Geistes- und Kulturwissenschaften eine neue Erfahrung, ebenso vielleicht wie die neuen Arten des Zusammenarbeitens, die die neuen digitalen Methoden erfordern. Daher stellen die digitalen Geistes- und Kulturwissenschaften eine große Chance, aber auch Herausforderung dar, und der Wandel, den sie vorschlagen, muss intensiv vorbereitet und diskutiert werden.

Dieser Beitrag behandelt einen Aspekt dieses Wandels, den Aufbau von digitalen Forschungsinfrastrukturen in den Geistes- und Kulturwissenschaften, die die neue Art des Zusammenarbeitens befruchten sollen. DARIAH-EU versucht eine nachhaltige Forschungsinfrastruktur zu etablieren, wie wir im nächsten Abschnitt beschreiben, bevor wir näher auf digitale Geisteswissenschaften eingehen und vor allem auf die Frage, was mit den großen Datenmengen in den Geistes- und Kulturwissenschaften geschehen kann. DARIAH-EU versucht, sich direkt an Communities auf Practice zu orientieren, die mit geistes- und kulturwissenschaftlichen Daten arbeiten wollen.

Vom Projekt zur nachhaltigen europäischen Forschungsinfrastruktur

DARIAH-EU ist eines von 48 Projekten auf der Roadmap des Europäischen Strategieforums für Forschungsinfrastrukturen (European Strategy Forum on Research Infrastructures, ESFRI). Nur fünf Projekte aus den Geistes- und Sozialwissenschaften werden durch ESFRI unterstützt, die Mehrheit der Projekte verteilt sich auf Natur- und Lebenswissenschaften sowie die technischen Disziplinen. DARIAH-EU ist eines der fünf sozial- und geisteswissenschaftlichen Projekte. Aus welchen Beweggründen heraus ist ESFRI entstanden, was hat es damit auf sich und welche Bedeutung hat ESFRI für DARIAH-EU?

ESFRI wurde 2002 als strategisches Instrument von der Europäischen Kommission initiiert, um die Entwicklung der europäischen Forschungslandschaft voranzutreiben und diese im internationalen Wettbewerb zu stärken.

Dabei geht es um die europaweite Koordination von Forschungsgrößgeräten, z.B. Riesenteleskopen² oder Dingen, die European Spallation Source³ heißen. Europas Willen und Fähigkeit, auf solch großer Ebene zusammenzuarbeiten, ist weltweit einmalig und gehört bewundert. Erklärtes Ziel von ESFRI ist, erstklassige Bedingungen in der europäischen Forschungslandschaft zu schaffen, um hier im globalen Maßstab herausragende Forschung zu ermöglichen.

Aufgabe von ESFRI ist es, einen kohärenten und strategiegeleiteten Ansatz für Forschungsinfrastrukturen in Europa zu unterstützen sowie multilaterale Initiativen, die zur besseren Nutzung und Entwicklung von Forschungsinfrastrukturen führen, sowohl auf europäischer als auch auf globaler Ebene zu fördern.⁴

Wissenschaftlicher Fortschritt ist ohne Spitzentechnologie inzwischen undenkbar. Die Forschungsmethoden und -prozesse sind zum Teil derart komplex und finanziell aufwendig, dass die Voraussetzungen dafür kaum ein Land allein stemmen könnte. In der Vergangenheit war es schwierig, länderübergreifende Forschungsinfrastrukturen aufzusetzen, da der adäquate rechtliche Rahmen fehlte, um entsprechend passende Partnerschaften ins Leben zu rufen.⁵

Zu den Errungenschaften von ESFRI zählt die Entwicklung einer Rechtsform für Europäische Forschungsinfrastrukturkonsortien (European Research Infrastructure Consortium, ERIC), um den europäischen Forschungsverbänden eine stabile Organisations- und Rechtsform zu geben, die von allen EU-Mitgliedsstaaten anerkannt ist. Das ERIC ist somit ein juristisches Instrument, was erst im Jahr 2009 von der Europäischen Kommission verabschiedet wurde. So kompliziert und aufwendig der initiale Abstimmungsprozess für ein ERIC auch ist, er ist mit einigen außerordentlichen Vorteilen für den dauerhaften Betrieb und die langfristig gesicherte Finanzierung von Forschungsinfrastrukturen verbunden.

Um die Nachhaltigkeit einer paneuropäischen Forschungsinfrastruktur zu gewährleisten, bedarf es einigen Vorlaufs, bis die Forschungsinfrastruktur schließlich Gestalt annimmt und nutzbar ist. ESFRI unterstützt dies durch

2 <http://www.eso.org/public/teles-instr/e-elt.html>

3 <http://www.ess-europe.org/>

4 Siehe http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri.

5 Siehe http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=eric1.

einen mehrstufigen Prozess. Vollständig von der Europäischen Kommission finanziell gefördert wird eine Vorbereitungsphase (preparatory phase), während der die Forschungsinfrastruktur sorgfältig geplant wird, sowohl in technischer als auch in administrativer und finanzieller Hinsicht. Idealerweise schließt die Vorbereitungsphase mit der Einreichung der ERIC-Statuten bei der Europäischen Kommission ab, bevor es direkt in die Aufbauphase (construction phase) übergeht. Die Praxis hat gezeigt, dass hier in der Regel eine Übergangsphase (transition phase) eingeschoben werden muss, um die rechtlichen Schritte für das Aufsetzen eines ERIC bis ins letzte Detail vorbereiten zu können.

So geschehen auch bei DARIAH-EU. Mehrere Treffen auf Ministerialebene aller zukünftigen Gründungsmitgliedsländer sowie zahlreiche Rücksprachen mit den entsprechenden Verantwortlichen der Europäischen Kommission waren notwendig, um sich auf eine finale Fassung der DARIAH-EU-ERIC-Statuten zu einigen, bevor sie im Herbst 2012 bei der Europäischen Kommission zur Begutachtung eingereicht werden konnte. Am Zustandekommen der DARIAH-EU-ERIC-Statuten waren zehn Nationen beteiligt: Deutschland, Dänemark, Frankreich, Griechenland, Irland, Kroatien, Niederlande, Serbien, Slowenien und Österreich.

An der Etablierung der Kommunikation zwischen den DARIAH-EU-Partnern sowohl auf Ministerialebene als auch auf operativer Ebene hatte die Abteilung Forschung & Entwicklung der SUB Göttingen einen maßgeblichen Anteil. Ohne den deutschen Beitrag würde DARIAH-EU nicht in seiner heutigen Form existieren. Zunächst mussten die richtigen Ansprechpartner gefunden werden. Das war gar keine so triviale Aufgabe angesichts der Tatsache, dass in jedem EU-Mitgliedsstaat die Forschungsförderung anders organisiert ist und der ERIC-Prozess noch in den Kinderschuhen steckt. Ebenso wichtig war es, die kulturellen Unterschiede, Bedürfnisse und verschiedenen Vorgehensweisen aufzufangen und in einen integrierten Prozess zu geleiten und letztlich passendes Lobbying für den Aufbau der digitalen Forschungsinfrastruktur zu betreiben. DARIAH-EU profitiert nicht nur von der von ESFRI entwickelten Rechtsform für Forschungsinfrastrukturen, sondern DARIAH-EU engagiert sich auch bei der Weiterentwicklung des Strategieforums, so z.B. in der ESFRI Implementation Working Group. Hauptaufgabe dieser Arbeitsgruppe ist die Unterstützung der ESFRI-Forschungsinfrastrukturen bei der Etablierung als europäische Rechtspersonen. Sie begleitet evaluierend den Prozess der Entwicklung einer nachhaltigen Rechtsform für

Forschungsinfrastrukturen und harmonisiert zugleich die bei diesem Prozess zutage tretenden Engpässe.

An die ca. sechs Jahre dauernde Aufbauphase schließt sich endlich die zeitlich unbefristete Nutzungsphase (operational phase) an. Wobei man den Aufbau und Betrieb einer digitalen Forschungsinfrastruktur nicht so strikt voneinander trennen kann wie das beispielsweise bei der Konstruktion und Nutzung von technischen Großgeräten der Fall ist, da die Anforderungen an eine solche digitale Forschungsinfrastruktur laufenden, aus der Forschung heraus geforderten Änderungen unterliegen, die berücksichtigt werden müssen. Nur durch dieses Wechselspiel von gleichzeitiger Nutzung und bedarfsgerechter Weiterentwicklung kann es gelingen, erstklassige und wegweisende Forschungsergebnisse mithilfe der digitalen Forschungsinfrastruktur zu gewinnen.

Digitale Geisteswissenschaften

DARIAH-EU selbst gehört zur Familie der digitalen Forschungsinfrastrukturen und wird sich den digitalen Methoden in den Geistes- und Kulturwissenschaften widmen sowie in den verwandten, d.h. qualitativen Sozialwissenschaften. DARIAH-EU ist eine Forschungsinfrastruktur, um die digitalen Geistes- und Kulturwissenschaften zu unterstützen. Sobald man so etwas sagt, wird man schnell mit der Frage konfrontiert, ob Geisteswissenschaften überhaupt Computer benötigen. Die Abneigungen sind im Zweifelsfall gegenseitig. Im Allgemeinen denkt der Computer nicht gern und Literatur schreibt er auch nur schlechte. Sogar Tim Berners Lee sieht das so: “there will be information, such as poetry and art, which requires the full human intellect for an understanding which will always be subjective” (Berners-Lee 1999).

Kann es das überhaupt geben: Digital Humanities? Wir glauben, niemand, der zu diesem Band beiträgt, wird bezweifeln, dass digitalen Methoden den Geisteswissenschaften nützen können. Praktische Beispiele gibt es genug: Zum Beispiel arbeiten wir seit Langem am King’s College London an der Digitalisierung der Stormont Papers, den Aufzeichnungen des nord-irischen Parlaments und der Troubles⁶. Wir hatten auch einmal ein Projekt mit dem

⁶ <http://stormontpapers.ahds.ac.uk/stormontpapers/index.html>

Englischen MetOffice⁷, nautische Logbücher des Ersten Weltkriegs zu analysieren. Diese können mit genauen Wetteraufzeichnungen rund um den Globus aufwarten und so der Analyse des Klimawandels dienen. Ohne Frage auch relevant, aber sehr schwierig zu digitalisieren, sind die Patientenakten des Victorianischen London. Leider verschwinden die physischen Archive hier immer mehr, weil den Londoner Krankenhäusern der Platz ausgeht. Und so wäre es wirklich wichtig, die Digitalisierung dieser Akten zu beschleunigen, um diese medizinhistorisch wichtigen und einmaligen Daten nicht zu verlieren.

Big Data in den Geistes- und Kulturwissenschaften

Und weil es so viele Menschen gibt auf der Welt und schon einige Geschichte, sprechen wir hier durchaus von Big Data. Ein passender Vergleich ist z.B. der eines der größten E-Science-Projekte mit zwei nicht untypischen geisteswissenschaftlichen Datensätzen. Die Sloan Digital Sky Survey⁸ hat weltweit Teleskop-Daten für die astronomische Forschung zusammengetragen und ist so Ende 2010 bei ungefähr 100 TB angekommen (vgl. Hand 2011).⁹ Wir können das nachfolgend mit zwei typischen Kollektionen in den Geisteswissenschaften vergleichen.

Die Spielberg-Sammlung von Holocaust Survivor Testimonials sind Video-Aufzeichnungen von Holocaust-Überlebenden mit ungefähr 200 TB an Daten. Diese Größenverhältnisse sind nicht überraschend, da diese Archive Videos und andere Multimedia-Dateien beinhalten. Auch in anderen Big-Data-Bereichen sind es gerade diese Art von Dateien, die die Größenverhältnisse dominieren.

Solche Datenmengen sind mittlerweile auch der renommierten Zeitschrift *Nature* aufgefallen, das 2010 ein Feature gebracht hat, das sich mit Big Data in den Geisteswissenschaften auseinandergesetzt hat (vgl. ebd.) – vor allem mit dem Culturomics-Projekt¹⁰, das anhand von Google-Daten die kulturelle Entwicklung der Sprache abzubilden sucht. Nach deren Visualisierung sind

7 <http://www.metoffice.gov.uk/>

8 <http://www.sdss.org/>

9 http://www.rochester.edu/data-science/assets/pdf/forum_slides/2012/Heidi%20Newberg%20slides.pdf

10 <http://www.culturomics.org/>

Geistes- und Kulturwissenschaften irgendwo in der Mitte der Big Data anzufinden.

Die Frage der Forschungsinfrastrukturen ist also nicht, ob, sondern wie man das geisteswissenschaftliche Forschungsmaterial zusammenbringt. In DARIAH-EU glauben wir, dass man sich hier auf das Arbeiten mit dem Material konzentrieren muss. DARIAH-EU baut auf der These auf, dass die Art und Weise des Arbeitens mit solchem digitalen Material nicht unbedingt so unterschiedlich sein muss, wie die Namen von immer neuen Subdisziplinen der Geisteswissenschaften zu suggerieren vermögen, weil es eben um digitale Methoden geht und wir daher beschränkt sind auf das, was Computer zu leisten vermögen.

Es geht also bei Projekten wie DARIAH-EU auch darum, neue Formen des Zusammenarbeitens zwischen Informatik und Geisteswissenschaften zu schaffen. Die Geisteswissenschaften (wie übrigens alle Wissenschaften) sind schon immer Meister der interdisziplinären Arbeit gewesen. Im Englischen hat man dafür mit Communities of Practice einen schönen Begriff gefunden: also – frei übersetzt – Leute, die gemeinsam etwas machen wollen oder gemeinsam arbeiten wollen. An diesen werden sich Projekte zur digitalen Forschungsinfrastruktur orientieren müssen und ihnen helfen, eigene neue Methoden zu entwickeln.

Eine entsprechende Forschungsinfrastruktur ist eine Herausforderung und eine eigene Wissenschaft für sich. Ein Vergleich zu klassischen technischen Infrastrukturen ist nicht einfach. Eine digitale Forschungsinfrastruktur ist keine Straße oder Autobahn, weil sie nie fertig ist, sondern sich immer nur von einem „stable release“ zur nächsten Überraschung weiterarbeiten muss. Man muss also organisatorisch flexibel sein und keine Straßen zementieren, sondern Dienste um Communities herum entwickeln.

Um diese Community-Dienste einzubetten, benötigt man ein flexibles organisatorisches Modell. DARIAH-EU hat ein solches Organisationsmodell entwickelt, das gleichzeitig zur paneuropäischen Zusammenarbeit verpflichtet. Das Kernstück von DARIAH-EU bilden die virtuellen Kompetenzzentren (Virtual Competency Centres, VCCs). Jedes ist an Themen ausgerichtet, die unserer langjährigen Erfahrung nach zur Unterstützung der digitalen Geistes- und Kulturwissenschaften unbedingt dazugehören.

Da ist einerseits das VCC e-Infrastructure, sozusagen das technische Herz von DARIAH-EU. Die Abteilung Forschung & Entwicklung der SUB Göttingen hat während der Vorbereitungsphase wesentlich zur Konzeption der technischen Basisinfrastruktur beigetragen, z.B. wurden erste Versionen des

Metadaten schemas und der technischen Architektur verfasst und fachspezifische Demonstratoren entwickelt, dies alles natürlich in enger Abstimmung mit Fachwissenschaftlern.

Ferner möchten wir die Wissenschaftler direkt in die Forschungsinfrastruktur einbinden mit dem VCC Research and Education. Wir glauben, dies ist einmalig in den digitalen Forschungsinfrastrukturen. Wir garantieren dadurch, dass DARIAH-EU eine Forschungsinfrastruktur von Fachwissenschaftlern für Fachwissenschaftler sein wird. Dies schließt auch die Ausbildung von Nachwuchswissenschaftlern in den digitalen Geisteswissenschaften ein, wofür DARIAH-EU Curricula auf europäischer Ebene abstimmt und weiterentwickelt, aber auch verschiedene Experten- und Methodenworkshops konzipiert und durchführt.

Schließlich unterstützen wir langlebige Wissenschaftsdaten mit dem VCC Scholarly Research Data. Interoperabilität, Nachnutzbarkeit, Langlebigkeit der Daten sowie Vertrauen in ihre Qualität sind die zentralen Themen zwischen fächerübergreifender Perspektive und fachlicher Spezifik. Die Langzeitarchivierung und -verfügbarkeit sowie Interpretierbarkeit von Forschungsdaten und Metadaten ist dabei Grundvoraussetzung für eine funktionierende Forschungsinfrastruktur. Die Einhaltung und Weiterentwicklung von Standards und die Einrichtung einer Qualitätskontrolle garantieren die Verlässlichkeit und Vertrauenswürdigkeit der registrierten Forschungsdaten.

Zuletzt darf auch ein VCC Impact, Outreach and Advocacy nicht fehlen, das die Schnittstelle zu einflussreichen Akteuren im weitesten Sinne ist, die DARIAH-EU politisch bzw. organisatorisch stärken und sichtbar machen können. Dem aktuellen Trend zur Messung der Aufwand/Nutzen-Relation im Wissenschaftsbetrieb folgend, erhebt es außerdem den Einfluss und misst den Mehrwert von DARIAH-EU mittels quantitativer und qualitativer Methoden. Insbesondere die Kooperationen und der Wissenstransfer zwischen geistes- und kulturwissenschaftlichen und anderen Disziplinen werden hierbei evaluiert.

Jedes VCC wird binational geführt: Je zwei Institutionen mit den notwendigen Kompetenzen aus verschiedenen Mitgliedsländern haben die Leitung eines VCCs inne und die verschiedenen Partnerinstitutionen erbringen ihren eigenen Beitrag zum vielfältigen Aufgabenspektrum der VCCs.

Die Gesamtorganisationsstruktur von DARIAH-EU zeigt Abbildung 1. Im Wesentlichen gibt es drei Gremien: die General Assembly, das Coordination Board und das Scientific Board, das DARIAH-EU in allen technischen und wissenschaftlichen Fragen berät. In der General Assembly sind

Vertreter der Finanzmittelgeber vertreten, das sind in den meisten Mitglieds-ländern Ministeriumsvertreter. Das Coordination Board ist die operative Ebene von DARIAH-EU, hier fließen alle Informationen zusammen. Es wird geleitet von einem der DARIAH-EU-Direktoren, die selbst das Board of Directors bilden. Dieses besteht zu Beginn aus drei Fachwissenschaftlern. Alle drei Direktoren engagieren sich jeweils mit einer halben Stelle in DARIAH-EU, mit ihren verbleibenden Stellenanteilen sind sie weiter aktiv in der Forschung tätig, um auf diese Weise so nah wie möglich an der aktuellen Forschung teilhaben zu können und die dort entstehenden Bedürfnisse direkt nach DARIAH-EU hineinbringen zu können. Dem DARIAH-EU Coordination Office (DCO) obliegt das gesamte Projektmanagement auf europäischer Ebene; es ist ebenso virtuell organisiert und agiert von Frankreich, den Niederlanden und Deutschland aus. Frankreich zeichnet für alle finanziellen und juristischen Entscheidungen verantwortlich, Deutschland übernimmt die all-gemeine Projektkoordination und Öffentlichkeitsarbeit, die Niederlande schließlich koordinieren die virtuellen Kompetenzzentren (Virtual Competency Centres, VCC).

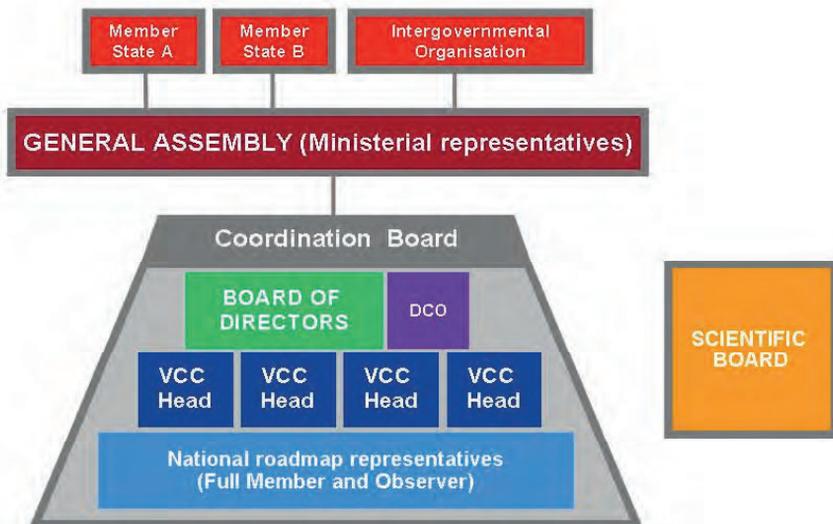


Abb. 1 Organisationsstruktur von DARIAH-EU

Communities of Practice

Auch technisch gesehen wird DARIAH-EU flexibel sein und sich an den Prinzipien der etablierten Webtechnologien und den Konzepten des Web 2.0 und Web 3.0 orientieren. So baut man effizient Dienste um Communities auf, die man auch gleich noch untereinander austauschen kann. Wir haben in unseren Kooperationsprojekten mit Wissenschafts-Communities gelernt, dass man viel von Innovationen von Infrastrukturen im Social-Web-Bereich lernen kann. Von diesen Kooperationsprojekten werden wir nun noch zwei Communities of Practice vorstellen, um die herum wir angefangen haben, Dienste aufzusetzen. DARIAH-EU selbst wird dann eine Art virtueller sozialer Marktplatz sein, um diese Dienste miteinander auszutauschen. Dies haben wir im Detail in Blanke et al. (2011) beschrieben.

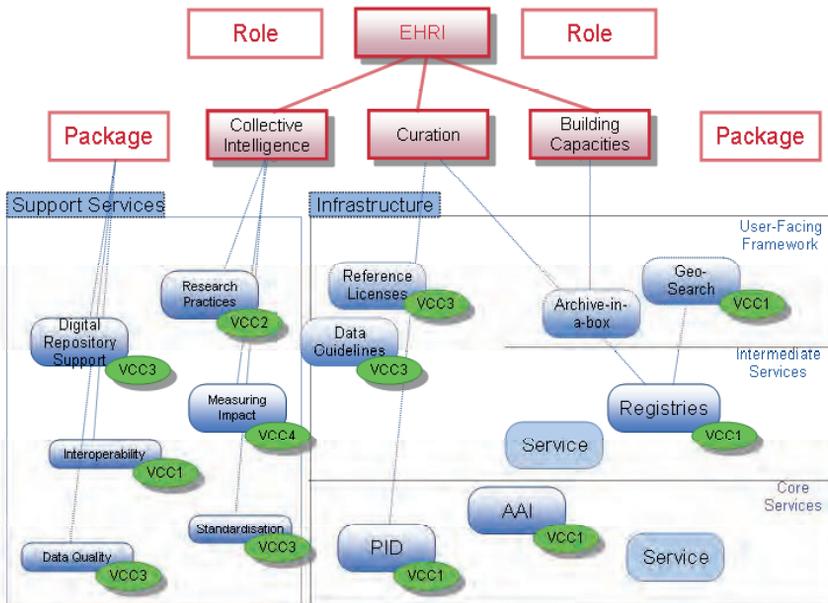


Abb. 2 Abbildung einiger DARIAH-EU-Dienste auf die Bedürfnisse von EHRI

Die European Holocaust Research Infrastructure (EHRI) ist Thema eines eigenen Beitrags in dieser Festschrift. Deswegen wollen wir uns in diesem Beitrag nur auf die wissenschaftlich-technische Kooperation mit DARIAH-EU konzentrieren. Abbildung 2 zeigt, wie einige DARIAH-EU-Dienste auf

die Bedürfnisse von EHRI abgebildet werden. Man kann hier sehen, wie und wo wir zusammenarbeiten wollen, z.B. bei der Data Curation oder mit speziell ausgerichteten digitalen Archiven.

Für DARIAH-EU ist EHRI auch deshalb so aufregend, weil es eben um mehr geht als um die Sonderwünsche einer zugegebenermaßen sehr besonderen und kleinen Forschungsgemeinschaft. Es geht darum, sich damit auseinanderzusetzen, wie der traditionelle Arbeitsplatz eines Geisteswissenschaftlers – das Archiv – in eine Virtuelle Forschungsumgebung übersetzt werden kann. Es geht um Virtuelle Forschungsumgebungen für dokumentenzentrierte Forschung, um traditionelle Archivstrukturen und neue Forschungsansätze.

Weil es aber um dokumentenzentrierte Forschung geht, können wir bei DARIAH-EU auf den Erfahrungsschatz einer anderen Community of Practice zurückgreifen. Viel mehr Erfahrung – jedenfalls jetzt noch – haben wir mit dieser zweiten Community of Practice, die wir Ihnen nun noch als letztes vorstellen: Textual Scholarship. Hier spielt das von der Abteilung Forschung & Entwicklung der SUB Göttingen geleitete Projekt TextGrid natürlich eine herausragende Rolle. DARIAH-EU und andere befreundete Organisationen sind hier schon jetzt, wie uns die Europäische Kommission gerne hätte. Erklärtes Ziel von DARIAH-EU ist es, „integrating activity“ zu sein.

TextGrid ist ebenfalls Thema eines anderen Beitrags in dieser Festschrift. TextGrid hat sich an den Anstrengungen einer existierenden Community orientiert und hat gleichzeitig die Technologie weiterentwickelt. Es ist eines der erfolgreichsten Projekte im E-Science-Bereich, das sich als erstes mit Interaktion und Annotation in Grid- und Cloud-Umgebungen auseinandergesetzt hat. Weil es eben jedoch nicht primär ein rein technisches Projekt ist, ist es gleichzeitig eingebettet in eine größere Familie von Projekten und dadurch Inspirationsquelle für neue Familienmitglieder. Wir geben Ihnen nur zwei Beispiele von Projekten aus der DARIAH-EU-Familie: Zum einen die 1641 Depositions¹¹, bei denen es um die Analyse einer irischen Rebellion gegen die englische Administration geht, sowie TEXTvire¹².

Man kann sogar sagen, dass die TextGrid-Zusammenarbeit Voraussetzung für DARIAH-EU ist. TEXTvire ist unser eigener TextGrid-Ableger am King's College London, das versucht, TextGrid um zweierlei Dinge zu erweitern. Zum einen versuchen wir, TextGrid einer „fremden Umgebung“

11 <http://1641.tcd.ie/>

12 <http://textvire.cerch.kcl.ac.uk/>

gerufen hat, um weiteren Terror zu legitimieren. Abbildung 3 ist eine Visualisierung der „1641 Depositions“ in IBMs ManyEyes-Programm. Es stellt Beziehungen zwischen Begriffen und Phrasen dar, die typisch für die Zeugenberichte sind.

Zusammenfassung

Dieser Beitrag hat sich mit dem Aufbau von DARIAH-EU beschäftigt, einer neuen multi-nationalen Forschungsinfrastruktur für die Geistes- und Kulturwissenschaften. Wir haben zunächst dargestellt, dass der Nachhaltigkeitsgedanke bei DARIAH-EU eine besondere Rolle spielt, was mit den ESFRI-Programm und dem Aufbau einer multinationalen rechtlichen Form namens ERIC zu tun hat. Von beiden erhoffen wir uns für DARIAH-EU eine Ausdauer, die über die übliche Projektlänge hinausgeht. Die Voraussetzungen dafür sind gut, wenn man berücksichtigt, dass sich nicht nur ein Land, sondern gleich mehrere Länder für DARIAH-EU verantwortlich erklären. Danach haben wir dargestellt, inwiefern dies besonders in den Geistes- und Kulturwissenschaften nötig ist, die mit der Ankunft von Big-Data-Mengen vor ganz neuen Herausforderungen stehen. Neue Methoden müssen nachhaltig entwickelt werden, aber auch neue Arten des Zusammenarbeitens.

DARIAH-EU sieht seine Aufgabe darin, um Communities of Practice neue Dienste und Methoden der Analyse, aber auch des Management von Daten zu entwickeln. Dabei wird es sich von seinen Erfahrungen in der Zusammenarbeit mit großen europäischen Projekten wie EHRI, aber auch mit nationalen Kollaborationen wie TextGrid, leiten lassen. DARIAH-EUs Aufgabe wird es sein, diese Erfahrungen auch anderen zukommen zu lassen und die Dienste, die aus diesen Projekten entstehen, auf eine europäische Ebene skalieren zu lassen. DARIAH-EU steht dabei erst am Anfang. Wir erwarten für 2013 die Verabschiedung der ERIC-Statuten, wobei schon jetzt die VCCs zusammenarbeiten. Aber erst, wenn das ERIC eingerichtet worden ist, kann man wirklich von einer nachhaltigen Entwicklung ausgehen, die auch für zukünftige Entwicklungen der europäischen Forschung relevant sein wird.

Literaturverzeichnis

- Berners-Lee, Tim (1999): Realising the Full Potential of the Web. In: *Technical Communication: Journal of the Society for Technical Communication* 46, no. 1: 79–82.
- Blanke, Tobias; Michael Bryant; Mark Hedges; Andreas Aschenbrenner; Michael Priddy (2011): Preparing DARIAH. In: *2011 IEEE 7th International Conference on e-Science*, Stockholm, Sweden, 5–8 December 2011. Piscataway, NJ: IEEE, pp. 158–165.
- Hand, Eric (2011): Culturomics: Word play. In: *Nature* 474, no. 7352: 436–440.

Forschungsdaten-Management

von Jens Klump und Jens Ludwig

Blick zurück: Entwicklung des Forschungsdaten-Managements in den letzten zehn Jahren

Das Deutsche GeoForschungsZentrum (GFZ) in Potsdam und das Zentrum für marine Umweltwissenschaften (MARUM) der Universität Bremen hatten nach erfolgreichen Pilotstudien 2004 zusammen mit der Technischen Informationsbibliothek Hannover (TIB) und dem Deutschen Klimarechenzentrum (DKRZ) das DFG-Projekt „Publikation und Zitierbarkeit wissenschaftlicher Primärdaten“ (STD-DOI) begonnen. Mit den bereits gesammelten Erfahrungen im Management geowissenschaftlicher Datenbestände am GFZ und am MARUM wollte man dieses Wissen mit der Expertise ausgewiesener Bibliotheken in einem Leistungszentrum für Forschungsinformation in den Geowissenschaften zusammenführen, dem „Zentrum für geowissenschaftliches Informationsmanagement“ (CeGIM). Am von der DFG geförderten Pilotprojekt waren das GFZ, das MARUM, die SUB Göttingen, die UB der TU Bergakademie Freiberg und die Bibliothek des Wissenschaftsparks Albert Einstein in Potsdam beteiligt. Dieses Projekt war der Ausgangspunkt der seitdem andauernden Zusammenarbeit der Abteilung Forschung & Entwicklung der SUB Göttingen und des Zentrums für Geoinformationstechnologie des GFZ bzw. seiner Vorgängereinheit, dem Daten- und Rechenzentrum des GFZ.

Schon sehr früh hatte man in den Geowissenschaften erkannt, dass Strukturen für die Bereitstellung und Langzeiterhaltung der Daten notwendig sind. Bereits für das Internationale Geophysikalische Jahr (IGY, 1957 bis 1958) wurden die World Data Center (WDC) eingerichtet, um die Daten des IGY bereitzustellen und zu erhalten (vgl. Dittert/Diepenbroek/Grobe 2001; Pfeiffenberger 2007). Darüber hinaus sollte das System der WDC in den Zeiten des Kalten Kriegs helfen, durch Zusammenarbeit und Transparenz zum Vertrauen zwischen Wissenschaftlern der beiden Machtblöcke beizutragen und dadurch zur politischen Entspannung beitragen. Mit dem Ende des „Kalten Krieges“ und der Entstehung des Internets verlor das System der

WDC in der bis dahin bestehenden Form seinen Sinn, sodass 2008 der International Council for Science (ICSU) beschloss, das System der WDC zu reformieren und in das World Data System (WDS) zu überführen.

Das entstehende Internet erlaubte einen immer einfacheren und schnelleren Austausch von Forschungsdaten, sodass bald die kulturelle Weiterentwicklung des Umgangs mit Forschungsdaten hinter den technischen Möglichkeiten zurückblieb (vgl. Klump et al. 2006). Ein Ergebnis der Unzufriedenheit mit dieser Entwicklung war die Veröffentlichung der „Berliner Erklärung über offenen Zugang zu wissenschaftlichem Wissen“ vom 22. Oktober 2003 (Berlin Declaration 2003). Die „Berliner Erklärung“ wird zwar in erster Linie als Meilenstein der Open-Access-Bewegung gesehen, jedoch beinhaltet sie auch die Forderung nach einem offenem Zugang zu Forschungsdaten und geht damit über die „Regeln für eine gute wissenschaftliche Praxis“ der DFG (1998) hinaus, die in erster Linie die Beweissicherung zur Verhinderung und Prüfung wissenschaftlichen Fehlverhaltens behandeln (vgl. Klump et al. 2006). Das Potenzial, das im offenen Zugang zu wissenschaftlichem Wissen steckt, wurde auch von der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) erkannt, weshalb deren Rat 2006 eine Empfehlung über den Zugang zu Daten aus öffentlich geförderter Forschung verabschiedete (vgl. OECD 2009). Es wird üblicherweise von den Mitgliedsstaaten der OECD erwartet, dass diese eine Empfehlung des OECD-Rates in nationale Gesetzgebung überführen. Im Falle der Empfehlung zum Umgang mit Forschungsdaten wurde dieser Prozess auch weiter von der OECD begleitet. Ein Ergebnis ist, dass die DFG seit 2010 in ihrem Leitfaden für Antragsteller eine Erklärung verlangt, wie mit den im Projekt gewonnenen Forschungsdaten umgegangen werden soll (vgl. DFG 2010), und das Thema des Umgangs mit Forschungsdaten auch systematisch weiterentwickelt. Noch weiter geht der im Februar 2013 veröffentlichte Erlass der US-Bundesregierung (vgl. Holdren 2013), die alle US-Bundesbehörden dazu verpflichtet, Ergebnisse öffentlich geförderter Forschung – dazu zählen Literatur und Daten – binnen zwölf Monate nach Veröffentlichung der Arbeiten kostenfrei zugänglich zu machen.

In der Praxis ist die Entwicklung eines systematischen und auch offenen Umgangs mit Forschungsdaten unter den Fachdisziplinen ungleich ausgeprägt. Ein offener Umgang mit Forschungsdaten ist vor allem in den Feldern zu beobachten, in denen der einzelne Forscher von einer Zusammenarbeit mit anderen Gruppen in seinem Feld profitiert (vgl. Neuroth et al. 2012). Dies drückt sich auch in dem Ungleichgewicht aus, dass die Bereitschaft, Daten

nachzunutzen überwiegt gegenüber der Bereitschaft, Daten mit anderen zu teilen (vgl. Borgman 2012). Die fehlende Bereitschaft dazu liegt wenigstens teilweise darin begründet, dass es heute wenig Anreize für den Wissenschaftler gibt, die zusätzliche Arbeit, die für das Bereitstellen von Daten für andere notwendig wäre, zu leisten. Darüber hinaus muss jedoch auch untersucht werden, wie die Akteure und ihre Rollen im Umgang mit Forschungsdaten verteilt sind und an welchen Institutionen sie angesiedelt sind.

Eine besondere Bedeutung in der Entwicklung neuer Praktiken im Umgang mit Forschungsdaten kommt den Bibliotheken zu. Insbesondere an akademischen Einrichtungen sind sie Teil der Infrastruktur und haben dabei unter anderem die Funktion der Gedächtnisorganisation und des Informationsdienstleisters. Bibliotheken könnten somit eine Rolle bei der Übernahme von digitalen Forschungsergebnissen spielen, um ihren Vertrieb, ihre Veröffentlichung und ihre langfristige Aufbewahrung zu organisieren. In der Ausbildung des Personals kommt das Thema Forschungsdaten gerade erst an und die konzeptionelle und technische Entwicklung in diesem Feld ist immer noch so schnell, dass Lehrmaterialien binnen kurzer Zeit veralten.

Bibliotheken sind gegenüber den Fachwissenschaften „fachfremd“. Das heißt, dass sie digitale Forschungsdaten nur übernehmen können und begrenzt zu deren systematischen fachlichen Beschreibung beitragen können. Auch die Entwicklung und der Betrieb der informationstechnischen Infrastrukturen, wie sie hier in Zukunft benötigt werden, ist bisher kein typischer Tätigkeitsbereich einer Bibliothek. Insofern müssen in Zukunft möglicherweise auch kooperative Modelle der Zusammenarbeit von Bibliotheken, Rechenzentren und disziplinären Kompetenzzentren für Forschungsdaten entwickelt werden. Diese Fragen, wie der Umgang mit Forschungsdaten verbessert werden kann und welche Rolle dabei Bibliotheken spielen könnten, wollen wir in diesem Kapitel erörtern.

Die Gegenwart: Aufgaben und Zuständigkeiten im Forschungsdaten-Management

Auch wenn Forschungsdaten-Management kein neues Thema ist, so wird es derzeit stärker als je zuvor diskutiert. Im Bereich der Bibliotheken wird in dem Thema oftmals eine Chance gesehen, sich angesichts der unklaren Rolle der Bibliotheken in der Informationsinfrastruktur der Zukunft ein neues Thema zu erschließen. Manchmal erfolgt dies mit einer Selbstverständlich-

keit und vielleicht auch Naivität, die bei Forschungsinstitutionen, die sich schon lange um Forschungsdatenbestände kümmern, die Skepsis gegenüber der Popularität des Themas bekräftigt. Um zu verstehen, welche Institutionen welche Rolle im Forschungsdaten-Management sinnvoll spielen können, ist es notwendig, den unscharfen Begriff des Forschungsdaten-Managements genauer aufzuschlüsseln.

Es kommt vor, dass Vertreter von Bibliotheken behaupten, dass sie bereits sehr lange Forschungsdaten-Management betreiben. Diese überraschende Aussage wird verständlich, wenn man sich vor Augen führt, wie breit der Begriff Forschungsdaten verwendet wird. Oft wird darunter alles verstanden, was in der Forschung benutzt wird und digital vorliegt, sodass auch z.B. digitale Artikel oder digitalisierte Bücher und Nachlässe als Forschungsdaten gelten. Um die eigentlichen Besonderheiten des Forschungsdaten-Managements in den Blick zu bekommen, ist es aber hilfreich, Forschungsdaten in einem engeren Sinne als die Daten zu verstehen, die im Forschungsprozess für die Untersuchung eines Forschungsgegenstands benutzt werden oder wurden. Der paradigmatische Fall sind Daten aus Messungen, aber auch Ergebnisse aus Simulationen oder Umfragen können als typisch gelten.

Forschungsdaten werden nicht nur begrifflich mit Publikationen assoziiert, sondern auch viele Aktivitäten des Forschungsdaten-Managements hängen mit Publikationen oder Publizieren zusammen. Die bereits erwähnte Empfehlung Nummer 7 der „Regeln für eine gute wissenschaftliche Praxis“, dass „Primärdaten als Grundlagen für Veröffentlichungen [...] auf haltbaren und gesicherten Trägern in der Institution, wo sie entstanden sind, für zehn Jahre aufbewahrt werden [sollen]“ (DFG 1998: 12), begründet beispielhaft die Notwendigkeit des Forschungsdaten-Managements als Dokumentation für Publikationen. Es gibt auch noch weitere Fälle, in denen sich Forschungsdaten-Management aus einer Pflicht zur Dokumentationen begründet, bei denen aber nicht unbedingt die Forschung selbst, sondern auf ihr basierende Entscheidungen oder Produkte überprüfbar sein sollen, wie z.B. Patientenbehandlungen in der Medizin oder kommerzielle Produkte in der Wirtschaft.

Eine etwas anders ausgerichtete Art von Forschungsdaten-Management zielt nicht darauf, Forschungsdaten als Dokumentation für Publikationen zu erhalten, sondern Forschungsdaten als Publikationen zur Nachnutzung bereitzustellen. Dadurch, dass das Veröffentlichende von Daten jeder Art durch das Internet viel einfacher geworden ist, ist es in großem Maßstab möglich geworden, Forschungsdaten zu publizieren. Insbesondere bei aufwendig erzeugten Forschungsdaten soll eine höhere Effizienz des Wissenschaftssys-

tems dadurch erreicht werden, dass Daten für neue Forschung nachgenutzt werden.

Bibliotheken werden traditionell als Sammler und Bewahrer von Publikationen verstanden und vor diesem Hintergrund scheint es naheliegend, sie auch als Akteure im Bereich der beiden eben beschriebenen Arten von Forschungsdaten-Management zu sehen. Typische Fragestellungen sind dann, wie Forschungsdaten in die bestehenden Repositorien übernommen und in den Katalogen nachgewiesen werden können. Für den Anwendungsfall der Dokumentation kann die Bibliothek wahrscheinlich auch ihre traditionellen Kompetenzen nutzbringend anbringen und ein wichtiger Akteur sein. Da auf solche Daten allerdings nur sehr selten zugegriffen wird und sie nicht mehr unmittelbar produktiv genutzt werden, handelt es sich aber um keine besonders prestigeträchtige Aktivität. Im Fall der Forschungsdaten-Publikationen für die Nachnutzung wird zwar wahrscheinlich auch selten auf den einzelnen Datensatz zugegriffen, aber durch die Nutzung für neue Forschung erhält das Forschungsdaten-Management ein positiveres Ansehen.

Ob Bibliotheken dafür die geeignetsten Institutionen sind, ist allerdings zweifelhaft, denn das Management von Forschungsdaten als Publikation für die Nachnutzung ist eine sehr voraussetzungsreiche Aktivität. Es existieren nicht ohne Grund viele disziplinspezifische Forschungsdatenzentren, da eine hohe Disziplinverankerung und großes Fachwissen notwendig sind. Typische Aktivitäten sind die aufwendige Aufbereitung und Dokumentation von Datensätzen, die Qualitätskontrolle von Forschungsdaten, die Pflege langfristiger Zeitreihen, die Interaktion mit der Fach-Community und die Begleitung der Entwicklung der Disziplin, um z.B. über Veränderungen der Anforderungen oder der Terminologie auf dem Laufenden zu bleiben. In einigen Fachdisziplinen wird die Veröffentlichung von Forschungsdaten in ganz bestimmten Zentren vorausgesetzt, wie z.B. der Protein Data Bank. Damit ist nicht ausgeschlossen, dass auch Bibliotheken Forschungsdaten veröffentlichen und pflegen oder gar die Rolle eines Forschungsdatenzentrums übernehmen, aber es erfordert eine fachliche Spezialisierung, die wohl nur sehr wenige Institutionen vorweisen können. Der Unterschied drückt sich auch in der Zielgruppe des Forschungsdaten-Managements aus: Während die Dokumentation von Forschungsdaten nur erfordert, einen lokalen Auftrag zu erfüllen, meist auf dem kleinsten gemeinsamen Nenner verschiedener Disziplinen, geht man mit der Pflege publizierter Forschungsdaten eine Verpflichtung gegenüber einer globalen und ausdifferenzierten Fachdisziplin ein.

von Forschungsdatenzentren bietet auch aus diesem Grund disziplinspezifische Forschungswerkzeuge an, die z.B. durch eine standardkonforme Datenerzeugung sowohl den Wissenschaftlern während des Projekts die Arbeit erleichtern, als auch den Aufwand für die Publikation und für das langfristige Datenmanagement für Forschungsdatenzentren reduzieren.

Ob sich damit der Gesamtaufwand für das Forschungsdaten-Management reduziert, wird sich zeigen müssen. Es ist aber davon auszugehen, dass das Forschungsdaten-Management im Projekt die Akzeptanz für das Datenmanagement verbessert, da die Wissenschaftler unmittelbar einen Nutzen haben. Hinter dieser Akzeptanzfrage verbirgt sich eine grundsätzlichere Schwierigkeit für das Wissenschaftssystem in Bezug auf Forschungsdaten: Die Wahrnehmung ist häufig, dass die Datenerzeugung die eigentlichen Kosten sind und die Datenmanagementkosten nur etwas Zusätzliches sind, auf das man auch verzichten könnte. Wenn man es aber für den gesamten Nutzungszeitraum der Daten betrachtet, der deutlich über das erzeugende Projekt hinausgehen kann, dann sind sie allerdings genauso ein Teil der Nutzungskosten, wie es die Erzeugungskosten sind. Datenmanagement im Projekt kann vielleicht den Aufwand für die Datenpublikation und langfristige Datenarchivierung zu einem gewissen Grad in das Projekt verschieben, das die Daten erzeugt, und damit hoffentlich eine realistischere Wahrnehmung und Kalkulation des Aufwands befördern.

Die Rolle von Infrastruktur im Forschungsdaten-Management

Für das Forschungsdaten-Management als Dokumentation sind die Institutionen verantwortlich, an denen die Wissenschaftler arbeiten. Dies ergibt sich relativ klar daraus, dass es sich um die Aufbewahrung der Daten für Verantwortungszwecke handelt, und die Institutionen diejenigen sind, die die Rahmenbedingungen für die Arbeit der Wissenschaftler organisieren. Es handelt sich um eine relativ klar definierte und zu einem gewissen Grad standardisierbare Aufgabe, für die die lokalen Infrastruktureinrichtungen oder -abteilungen die richtigen Akteure sind. Für die Nachnutzung wiederum wurde oben ausgeführt, dass es sich um eine komplexe, disziplinspezifische und höchstens teilweise standardisierbare Aufgabe handelt, die wahrscheinlich am besten von überregionalen oder internationalen, spezialisierten und in den Disziplinen verankerten Zentren übernommen wird. Wer ist aber der richtige Akteur, um das Forschungsdaten-Management in Projekten zu übernehmen?

Dies kann sich je nach Kontext unterschiedlich darstellen. Wenn es sich um relativ technik-affine Disziplinen oder um geringe Aufwände handelt, dann übernehmen die Wissenschaftler das gegebenenfalls selbst – einige Disziplinen haben auch verwandte Zweige der Informatik, die in diesen Fällen kompetente Ansprechpartner sein können. Die zentralen Forschungsdatenzentren, die die Publikationen von Forschungsdaten und ihre langfristige Pflege übernehmen, bringen sicherlich ebenfalls die inhaltliche Kompetenz mit, würden aber als zentrale Einrichtung vor der Herausforderung stehen, sehr viele verteilte Forschungsprojekte unterstützen zu müssen. Zusätzlich zu dem Skalierungsproblem wäre es schwierig, in die jeweiligen Projektgegebenheiten eingebunden zu sein und ihren oftmals lokalen Besonderheiten gerecht zu werden, sodass Forschungsdatenzentren wahrscheinlich nur in Fällen besonderer Bedeutung Projekte direkt beim Forschungsdaten-Management unterstützen können (wie z. B. die langfristig angelegten Umfrageprojekte in den Sozialwissenschaften) und sich eher auf die Entwicklung zentraler Werkzeuge und Standards im Sinne von Referenz-Sammlungen (vgl. National Science Board 2005) konzentrieren.

Wie ist aber die Situation, wenn das Datenmanagement in Projekten nicht effizient von den Wissenschaftlern selbst übernommen werden kann, in interdisziplinären Projekten oder wenn keine fachbezogene Informatik ausgeprägt wurde oder lokal vorhanden ist? Im Prinzip handelt es sich um komplexe und sehr projektspezifische Aufgaben, die aber nicht unbedingt vollständig disziplinspezifisch sein müssen. Einrichtungen wie Bibliotheken oder Rechenzentren können dafür die richtigen Institutionen sein und kompetentes Personal haben, wenn sie die Tätigkeit des Forschungsdaten-Managements nicht nur unter dem Vorzeichen der Publikation betrachten, sondern sich auf die Unterstützung des Forschungsprozesses einlassen. Sie können direkt im Projekt integriert sein als „embedded data managers“ und für die gesamte Institution als lokale Forschungsdaten-Support-Teams dienen. Wenn sie über die Projekte, in denen sich diese Investition lohnt, finanziert sind, dann kann diese Form des Forschungsdaten-Managements auch skalieren.

Es ist klar, dass diese lokalen Support-Teams nicht die Fachkompetenz zentraler Forschungsdatenzentren ersetzen können, so wie die Zentren nicht die Integration im Projekt vor Ort ersetzen können. Die Support-Teams aber können wissen, wann sie welche spezialisierten Zentren hinzuziehen müssen. Es bietet sich hier die Möglichkeit für ein sinnvolles Zusammenspiel von Infrastruktureinrichtungen. Und erst dieses Zusammenspiel wird das sein können, was wir eigentlich Forschungsdaten-Infrastruktur nennen sollten.

Tabelle 1:

Übersicht über verschiedene Formen des Forschungsdaten-Managements

	Dokumentation für Publikationen	Nachnutzung als Publikation	Datenmanagement im Projekt
Ziel	Nachvollziehbarkeit für Verantwortungszwecke	bessere und effizientere Forschung durch Nutzung bereits vorhandener Daten	bessere und effizientere Forschung durch besseres Datenmanagement in Projekten
Auswahlkriterium	Daten, die Grundlage einer Publikation sind, oder entsprechend Auflagen	fachwissenschaftliche Qualität der Daten und Kosten-Nutzen-Abschätzung	Relevanz der Projekte und Kosten-Nutzen-Abschätzung
Preismaßstab, Kostenfaktor	Volumen	Komplexität oder Anzahl Datensätze	Aufwand, Abschätzung anhand von Personenmonaten
Zielgruppe	Universität/Institution (weniger Wissenschaftler)	Disziplin, Forschungs-Community	Forschergruppe/ Forschungsprojekt
Dienstleister	Lokale Service-Infrastruktureinrichtungen. z.B. Rechenzentren und Bibliotheken	spezialisierte Datenzentren mit disziplinärer Verankerung und Kompetenz	lokale Zentren mit Support-Teams für Forschungsdaten-Management
Aufbewahrungsdauer	lang, aber begrenzt (zehn Jahre oder entsprechend der Auflagen)	lang und unbestimmt (potenziell unbegrenzt)	kurz und begrenzt (Projektlaufzeit)
Zugriffshäufigkeit	sehr selten	selten	häufig
Volumen	hoch	mittel	mittel
Aufgabencharakteristik	standardisierbar, teilweise automatisierbar	komplex, begrenzt automatisierbar	je Projekt individuell angepasste Maßnahmen, personalintensiv, kaum automatisierbar

Bisher wurde in diesem Text der Begriff Infrastruktur nicht weiter erläutert – aber anhand der Behauptung, dass erst das Zusammenspiel von lokalen und zentralen Einrichtungen die Forschungsdaten-Infrastruktur darstellt, sol-

len zwei Aspekte hervorgehoben werden: 1. Infrastruktur ist nicht nur Technik; 2. Infrastruktur besteht nicht nur aus einzelnen Komponenten.

- Zu häufig wird explizit oder implizit ein technischer Service oder eine Technologie als die Lösung einer Aufgabe im Bereich des Forschungsdaten-Managements betrachtet, anstelle die Organisationen und persönlichen Dienstleistungen als das Fundamentale anzusehen. “When dealing with infrastructures, we need to look to the whole array of organizational forms, practices, and institutions which accompany, make possible, and inflect the development of new technology [...] People, routines, forms, and classification systems are as integral to information handling as computers, Ethernet cables, and Web protocols” (Edwards et al. 2007: 3). Oder – um einen pointierten Artikeltitle der California Digital Library zu zitieren: “Preservation Is Not a Place” (Abrams et al. 2009). Ein Repositorium alleine löst keine Aufgaben, sondern das wird nur durch Prozesse, in denen Personen Technik benutzen, erreicht. Und wenn das Ziel eine Infrastruktur für die öffentliche und langfristige Verfügbarkeit von Forschungsdaten ist, dann ist dies nur durch nachhaltige Organisations- und Personalressourcen zu erreichen, die die Bereitstellung und Pflege von Forschungsdaten fortwährend betreiben, obwohl es zu Veränderungen in allen Bereichen wie z.B. der benutzten Technologie, dem Hintergrundwissen der Nutzergruppe und ihren Erwartungen kommt. Zu den größten Herausforderungen im Bereich des Forschungsdaten-Managements zählt sicherlich dieser Umgang mit Veränderungen und die Einplanung eines eigenen Lebenszyklus für Infrastrukturkomponenten, die entwickelt, genutzt, irgendwann nicht mehr die funktionellen oder ökonomischen Anforderungen erfüllen und erneuert, abgewickelt oder ersetzt werden müssen.
- Zudem ist Infrastruktur nicht nur eine einzelne Einrichtung, ein einzelner Service oder eine einzelne Entwicklung. Einzelne Komponenten sind immer auf ihren Kontext ausgerichtet und erst im Zusammenspiel mit anderen Komponenten werden sie für die verschiedensten Situationen und Kontexte nutzbar. “[T]rue infrastructures only begin to form when locally constructed, centrally controlled systems are linked into networks and internetworks governed by distributed control and coordination processes” (Edwards et al. 2007: 7). Für das Forschungsdaten-Management ist dafür eine zentrale Frage, wie die verschiedenen Institutionen und Disziplinen zusammenarbeiten und wie disziplinspezifisch und wie generisch sowohl die einzelnen Dienste als auch die Institutionen sein müssen

und können. Und selbst die Disziplinen sind unscharfe Konzepte, die beliebig ausdifferenziert werden können. Betrachten wir die Physik als eigene Disziplin? Wenn wir sie als zu groß ansehen, wäre die Astrophysik oder erst die Radioastronomie hinreichend eingegrenzt? Oder sind die einzelnen Großprojekte schon so spezifisch geworden, dass man sie als eigenständig betrachten muss? Gleichzeitig finden wir im Datenmanagement nicht nur disziplinspezifische Anforderungen, sondern auch viele methodenspezifische Aufgaben, die sich in einer Vielzahl von Disziplinen finden und für die generische Dienste Synergiepotenzial versprechen.

Es existieren bereits eine Reihe von einzelnen Infrastrukturkomponenten, die sich in einigen Disziplinen oder Teildisziplinen gut zusammenfügen, und auch althergebrachte Netzwerk wie das oben beschriebene WDC oder WDS sowie neue Initiativen wie die Research Data Alliance¹. Es ist wahrscheinlich aber relativ unumstritten, dass es noch keine Forschungsdaten-Infrastruktur in dem Sinne gibt, wie man Infrastruktur gerne hätte: überall verfügbar, robust und transparent. Die hier beschriebenen Arten von Forschungsdaten-Management und die sich daraus ergebenden Anforderungen an Infrastruktureinrichtungen umreißen einen Rahmen, aber lassen noch viele Fragen unbeantwortet.

Mögliche Entwicklungen: Wie kann und wie sollte die Landschaft in zehn Jahren aussehen?

Die technische Entwicklung schreitet weiter schnell voran. Einiges, was uns heute schon selbstverständlich vorkommt, existierte vor zehn Jahren noch nicht einmal in unserer Vorstellung. Im Zusammenhang mit Forschungsdaten illustrieren die Begriffe „Grid“ und „Cloud“ gut die Dynamik dieser Entwicklung. Anfang der 2000er-Jahre kam der Begriff „Grid“ auf als das Versprechen, unbegrenzte IT-Ressourcen quasi „aus der Steckdose“ beziehen zu können. Für den privaten Nutzer schien das ohne Relevanz zu sein, da sich das Grid-Konzept in erster Linie an Nutzergemeinschaften orientierte. Gewerbliche Anwendungen, wie sie in der Förderung des BMBF auch gedacht waren, wurden nie etabliert, da die angesprochenen Firmen kein Vertrauen in die Sicherheit der Grid-Anwendungen hatten (vgl. Klump 2008).

¹ <http://rd-alliance.org>

Zeitgleich mit dem Aufkommen der Smartphones nach Vorbild des iPhone im Jahre 2007 und den neuen, damit verbundenen Online-Diensten kamen zwei neue Begriffe auf: Cloud Computing und App.

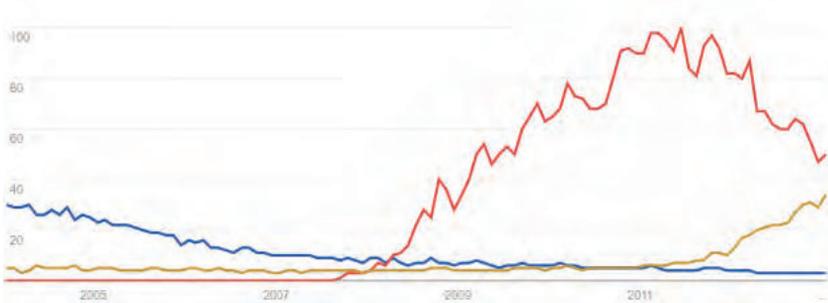


Abb. 2 Histogramm der Anfragen bei Google zu den Begriffen „Grid Computing“ (blau), „Cloud Computing“ (rot) und „Big Data“ (gelb) im Januar 2013.
Quelle: Google Trends

Mit dem Aufkommen des Begriffs „Cloud“ verschwand der Begriff „Grid“ praktisch in der Bedeutungslosigkeit (Abb. 2). Selbst der Begriff „Cloud“ ist unbedeutend verglichen mit dem Suchbegriff „App“, der sich als Synonym für Anwendungen (applications) auf mobilen Geräten etabliert hat (Abb. 3). Soziale Netzwerke spielten vor 2004 noch eine untergeordnete Rolle, heute hat Facebook rund eine Milliarde Nutzer.

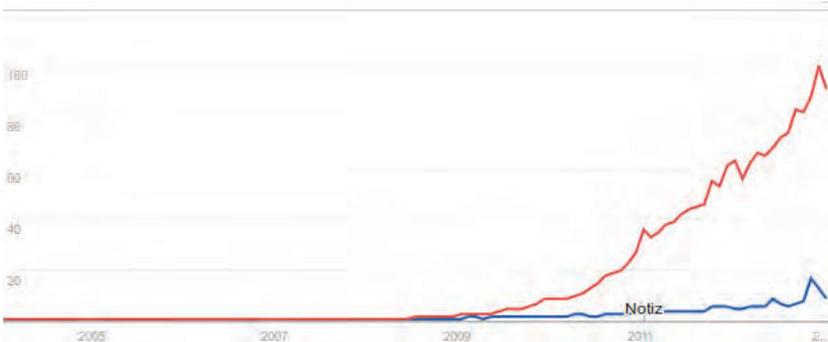


Abb. 3 Histogramm der Anfragen bei Google zu den Begriffen „Cloud Computing“ (blau) und „App“ (rot). Diese beiden Begriffe sind heute auch im Management von Forschungsdaten präsent, spielten aber noch vor 2010 fast keine Rolle.
Quelle: Google Trends

Was bedeutet das für Technologien und Dienste im Hinblick auf den Umgang mit Forschungsdaten? Ein Blick auf die oben skizzierten Trends zeigt die Dynamik der Entwicklung und macht deutlich, wie schwer es ist, die Entwicklung des Umgangs mit Forschungsdaten für die nächsten zehn Jahren vorherzusagen. Es ist unmöglich vorherzusagen, welche technischen Lösungen zur Verfügung stehen werden. Auch Trends lassen sich nur in begrenztem Maße identifizieren, denn die Entwicklung wird weiterhin stark von „disruptive innovation“-Mustern beeinflusst, was für sich selbst wiederum einen Trend in der weiteren Entwicklung darstellt.

Was ist nun wirklich ein Trend und was ist nur aufgebauscht? Ist „Big Data“ tatsächlich ein bedeutender Trend in der Wissenschaft? Werden die digital natives, die eine Welt ohne Internet nicht kennen, als die Wissenschaftler von morgen diese Technologien anders und freier nutzen? Hier hilft es, von den Technologien zu abstrahieren und zu fragen, welche Prozesse technisch unterstützt werden sollen.

Im Jahr 2003 erschienen der einflussreiche Beitrag „e-Science and its implications“ von Hey und Trefethen (2003), mit dem der Begriff der „Datenflut“ (data deluge) in die Diskussion eingeführt wurde. Die Diskussion darüber war damals noch sehr mit den erwarteten Datenmengen befasst. Mit dem technischen Fortschritt kamen jedoch auch andere Aspekte mit ins Blickfeld – nämlich die Möglichkeit, durch explorative Analyse der Daten neue Hypothesen zu formulieren und zu prüfen. Hier hängt der wissenschaftliche Fortschritt unmittelbar an der Verfügbarkeit der Daten und der Möglichkeit, sie zu verarbeiten – was auch mit dem Begriff „data intensive science“ bezeichnet wird (vgl. McNally et al. 2012). Der Informatik-Pionier Jim Gray sprach auch von einem Paradigmenwechsel (vgl. Hey/Tansley/Tolle 2009). Aber ist dieser Paradigmenwechsel eingetreten?

Nimmt man die bewilligten Projektmittel als Näherungswert für die Größe eines Projekts und das zu erwartete Datenvolumens, so zeigt sich, dass auch die Verteilung von Forschungsprojekten mit einer „long-tail economy“ (Anderson 2004) beschrieben werden kann. Nur relativ wenige Projekte erzeugen sehr große Datenmengen, die allermeisten Projekte erzeugen relativ kleine Datenmengen. Bei diesen Daten handelt es sich jedoch überwiegend um intellektuell vernetzte Objekte, die einen hohen Grad an semantischer und struktureller Komplexität besitzen und oft auch zwischen Projekten miteinander vernetzt sind (vgl. Heidorn 2008).

Eine weitere bedeutende Entwicklung ist die Entstehung sozialer Netzwerke und einer Vielzahl von Projekten mit Inhalten, die – teils anonym –

von Nutzern erstellt werden. Das herausragendste Beispiel von nutzergenerierten Inhalten ist die Online-Enzyklopädie Wikipedia. Mit dem Erscheinen dieser vielen Projekte wurde erwartet, dass dies auch Auswirkungen auf die Wissenschaft haben wird, insbesondere durch die „digital natives“, die eine Welt ohne Internet gar nicht kennen. Von ihnen wurde erwartet, dass sie die neuen Werkzeuge, wie zum Beispiel Wikis, für eine offenere wissenschaftliche Kommunikation nutzen. Der Bericht „Researchers of Tomorrow“, der von der British Library, JISC und HEFCE (2012) herausgegeben wurde, untersuchte durch Befragung von etwa 6000 graduierten Studenten, wie sich das Verhalten der „digital natives“ als Wissenschaftler von dem der vorangegangenen Generation ihrer wissenschaftlichen Betreuer unterscheidet. Zur Überraschung der Autoren stellte sich heraus, dass die „digital natives“ im Privatleben zwar die neuen Möglichkeiten des Internets und der Informationstechnologie intensiv nutzen, in ihrer wissenschaftlichen Praxis jedoch das Nutzerverhalten ihrer betreuenden Wissenschaftler kopierten, da sie diese als erfolgreiche Vorbilder ansehen. Es ist daher nicht zu erwarten, dass sich allein durch das Heranwachsen der „digital natives“ die Bereitschaft zur Zusammenarbeit im Netz, der Umgang mit Forschungsdaten, und die Bereitschaft, diese mit anderen zu teilen, rasch ändern – solange dies nicht als kulturell verankerte Wissenschaftspraxis vermittelt wird. Hinzu kommt, dass sich wissenschaftliche Arbeit hinsichtlich der Motivation der Autoren grundlegend von anderen „Open Initiatives“ unterscheidet (vgl. Klump 2012).

Herausforderungen an die Organisation

Die Anwendung der modernen Informations- und Kommunikationstechnik im Wissenschaftsbereich hat es deutlich einfacher gemacht, institutionsübergreifend zu kooperieren. Insgesamt nehmen kooperative Strukturen in der Forschung zu und der Umgang mit Forschungsdaten ist in den Bereichen bereits besser ausgebildet, in denen kooperative Strukturen vorherrschen (vgl. Neuroth et al. 2012). Der Zugewinn an Reputation durch vernetzte Projekte lässt Kooperation inzwischen oftmals als notwendige Arbeitsweise in der modernen Wissenschaft erscheinen. Je effizienter man einzelne Ergebnisse und technische Möglichkeiten ausnutzen will, desto flexibler und sachbezogener muss die Kooperation sein. Allerdings muss hier eine gewisse Spannung konstatiert werden: Der einzelne Wissenschaftler ist weiterhin an einer Organisation angestellt, die für ihn grundsätzliche Rahmenbedingungen be-

reitstellt, aber die Kooperation findet nicht mehr nur innerhalb einer Organisation statt.

Formen kooperativer Forschung stellen die Organisationen, in denen die daran teilnehmenden Wissenschaftler beheimatet sind, vor neue Herausforderungen. Es muss ein Rahmen geschaffen werden, der diese extremen Anforderungen an flexibler Kooperationen – geradezu volatiler, nur noch aufgabenbezogener Kooperation – erfüllen kann. Wünschenswert wäre, wenn Wissenschaftler sich allein mit ihrer speziellen Expertise in diese Projekte einbringen könnten, in einer Art Matrixorganisation. Dies umzusetzen ist jedoch auf der Ebene der Individuen schwierig, da Forschung nicht in Geschäftsprozessen denkt und auch die Struktur der Forschungsförderung und der Einwerbung von Mitarbeiterstellen durch Drittmittel diese Art von Organisation nicht vorsieht.

Die Anforderungen einer flexiblen „Matrixorganisation“ stehen in einem gewissen Widerspruch zu den Erwartungen der Heimatorganisationen, die bei der Zuteilung von Ressourcen fragen, was Kooperationen ihnen für Nutzen bringen. Konflikte entstehen dann notwendigerweise, wenn diese Anforderungen sich schneller ändern als die Ziele der beheimatenden Organisation. So fördert der technische Fortschritt zwar globale Kooperation, aber nicht notwendigerweise auch lokale. Die Dynamik der Forschung erfordert zum Teil spontane Organisation, wie sie durch neue Technologien möglich geworden ist – diese steht aber oft im Widerspruch zu lokaler Organisation, in der zum Beispiel eine Abteilung den Nutzen eines Projekts für eine übergeordnete Einrichtung demonstrieren muss, der für die Projektpartner evident ist.

Herausforderungen an die Infrastruktur

Infrastruktureinrichtungen stehen im Bezug auf die rasche Entwicklung der Informationstechnik und ihrer Nutzungsmöglichkeiten vor einem Grunddilemma: Es ist ihre Aufgabe, stabile Dienste zu betreiben und diese stetig zu verbessern. Für diese Situation einer „sustaining innovation“ sind Infrastruktureinrichtungen im Allgemeinen gut vorbereitet. Auch für diese Dienste ist es notwendig, ihren Lebenszyklus zu planen und sie an technische Innovationen anzupassen. Immer wieder tauchen jedoch neue Technologien auf, deren Bedeutung für die Arbeitsweise der Forschung nicht von Anfang an klar zu erkennen ist. Aufgrund ihrer Zielsetzung, stabile Dienste anzubieten,

fällt es Infrastruktureinrichtungen schwer, sich auf neue Entwicklungen einzulassen, deren langfristiger Erfolg noch nicht abzuschätzen ist. Im schlimmsten Fall kann dies dazu führen, dass sich eine Institution auf eine fundamentale Neuerung nicht einlassen kann und letztlich scheitert. Dieses Innovationsmuster wird „disruptive innovation“ genannt (vgl. Christensen 2003). Insbesondere wissenschaftliche Bibliotheken werden als gefährdet angesehen (vgl. Lewis 2004).

Zudem sind auch Neuerungen in der Informationstechnologie oft kurzlebig. Viele der, zum Beispiel, von Google angebotenen Dienste haben eine durchschnittliche Lebensdauer von etwa drei Jahren (vgl. Arthur 2013), was in etwa der Dauer von Innovationszyklen in der Informationstechnik entspricht. Das Beispiel Google zeigt, dass es auch für einen Weltkonzern nicht möglich ist, den Erfolg eines seiner Dienste vorherzusehen, und deshalb begegnet man dieser Herausforderung, in dem man ein Portfolio von Diensten auf einer gemeinsamen Plattform entwickelt. Einzelne Dienste können dann bei ausbleibendem Erfolg wieder abgeschaltet werden, ohne den Betrieb der Plattform als Ganzes zu beeinträchtigen.

Eine erfolgreiche Strategie für Infrastruktureinrichtungen könnte daher sein, ein modularisiertes Portfolio von Diensten zu entwickeln, das auf einer gemeinsamen Plattform aufbaut. Diese Strategie würde es den Einrichtungen erlauben, die Dienste flexibel den sich stets ändernden Bedürfnissen der Forschung anzupassen, während die darunter liegende Plattform stetig weiterentwickelt werden kann. Auf diese Weise ließe sich der Widerspruch zwischen dem Anspruch der Infrastruktur an Stabilität und der Anforderung nach flexiblen, möglicherweise kurzlebigen Anwendungen überbrücken.

Ein Maßstab, an dem neue Entwicklungen in der Informationstechnik gemessen werden können, ist, inwieweit sie die Prozesse der Forschung und Kommunikation unterstützen. Zu schnelle Änderungen in den technischen Rahmenbedingungen erschweren es potenziellen Nutzern, ihre Entwicklungen auf die neuen Dienste aufzubauen. Im Rahmen des D-Grid wurde, zum Beispiel, kritisiert, dass sich die Grid-Schnittstellen zu schnell änderten, um Anwendungen zu entwickeln, die sich in die Arbeitsabläufe der Forschung einpassen. Lieber hätte man mit älteren Versionen der Grid-Middleware gearbeitet, als die eigenen Anwendungen ständig auf neue Versionen der Grid-Middleware anpassen zu müssen.

Für Infrastruktureinrichtungen der akademischen Informationsversorgung ist die Einschätzung neuer Technologien schwer, da man sich dort nicht in der Lage sieht, technologische „Versuchsballons“ zu starten und gegebenen-

falls auch nach kurzer Zeit wieder aufzugeben. Zudem wurden Informationsinfrastrukturen meist als monolithische, in sich geschlossene Anwendungen gebaut. Um eine bessere Anpassung der Dienste an die Bedürfnisse der Forschung zu erreichen, wurden meist Nutzerbeiräte eingerichtet, um einen direkteren Austausch mit den Nutzern zu erreichen. Doch auch etablierte Nutzergruppen sind vor den Herausforderungen einer „disruptive innovation“ nicht gefeit (vgl. Christensen 2003).

Literaturverzeichnis

- Abrams, S.; P. Cruse; J. Kunze (2009): Preservation Is Not a Place. In: *International Journal of Digital Curation*, 4 (1): 8–21, doi:10.2218/ijdc.v4i1.72.
- Anderson, C. (2004): The Long Tail, In: *Wired*. Online: http://www.wired.com/wired/archive/12.10/tail_pr.html.
- Arthur, C. (2013): Google Keep? It'll probably be with us until March 2017 – on average. In: *The Guardian*, 22. März 2013. Online: <http://www.guardian.co.uk/technology/2013/mar/22/google-keep-services-closed>.
- Berlin Declaration (2003): Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. <http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung/>.
- Borgman, C. L. (2012): The Conundrum of Sharing Research Data. In: *Journal of the American Society for Information Science and Technology*, 63 (6): 1059–1078, doi:10.1002/asi.22634.
- British Library; HEFC; JISC (2012): Researchers of Tomorrow – The research behaviour of Generation Y doctoral students. London: JISC. Online: <http://www.jisc.ac.uk/publications/reports/2012/researchers-of-tomorrow>.
- Christensen, C. M. (2003): *The Innovator's Dilemma*. New York, NY: HarperCollins Publishers.
- DFG (1998): *Sicherung guter wissenschaftlicher Praxis*. Bonn: Deutsche Forschungsgemeinschaft.
- DFG (2010): Merkblatt für Anträge auf Sachbeihilfen mit Leitfaden für die Antragstellung und ergänzenden Leitfäden für die Antragstellung für Projekte mit Verwertungspotenzial, für die Antragstellung für Projekte im Rahmen einer Kooperation mit Entwicklungsländern, Deutsche Forschungsgemeinschaft (DFG), Bonn. http://www.dfg.de/download/formulare/1_02/1_02.pdf.
- Dittert, N.; M. Diepenbroek; H. Grobe (2001): Scientific data must be made available to all. In: *Nature*, 414 (6862): 393, doi:10.1038/35106716.

- Edwards, P.; S. Jackson; G. Bowker; C. Knobel (2007): Understanding Infrastructure: Dynamics, Tensions, and Design. <http://hdl.handle.net/2027.42/49353>.
- Heidorn, P. B. (2008): Shedding Light on the Dark Data in the Long Tail of Science. In: *Lib. Trends*, 57 (2), 280–299, doi:10.1353/lib.0.0036.
- Hey, T.; S. Tansley; K. Tolle (Hrsg.) (2009): *The Fourth Paradigm: Data-Intensive Scientific Discovery* (v 1.1). Redmond, WA: Microsoft Research. Online: <http://research.microsoft.com/en-us/collaboration/fourthparadigm/>.
- Hey, T.; A. Trefethen (2003): e-Science and its implications. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 361 (1809): 1809–1825, doi:10.1098/rsta.2003.1224.
- Holdren, J. P. (2013): Increasing Access to the Results of Federally Funded Scientific Research. The White House, Office of Science and Technology Policy. http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.
- Klump, J. (2008): *Anforderungen von e-Science und Grid-Technologie an die Archivierung wissenschaftlicher Daten*, nestor-Materialien, Kompetenznetzwerk Langzeitarchivierung (nestor). Frankfurt (Main). <http://nbn-resolving.de/urn:nbn:de:0008-2008040103>.
- Klump, J. (2012): Offener Zugang zu Forschungsdaten: Open Data und Open Access to Data – Die ungleichen Geschwister. In: U. Herb (Hrsg.): *Open Initiatives: Offenheit in der digitalen Welt und Wissenschaft*. Saarbrücken: Universaar, S. 45–53. Online: <http://nbn-resolving.de/urn:nbn:de:bsz:291-universaar-873>.
- Klump, J.; R. Bertelmann; J. Brase; M. Diepenbroek; H. Grobe; H. Höck; M. Lautenschlager; U. Schindler; I. Sens; J. Wächter (2006): Data publication in the Open Access Initiative. In: *Data Science Journal*, 5, 79–83, doi:10.2481/dsj.5.79.
- Lewis, D. W. (2004): The Innovator’s Dilemma: Disruptive Change and Academic Libraries. In: *Library Administration & Management*, 18 (2), 68–74. Online: <http://hdl.handle.net/1805/173>.
- National Science Board (2005): Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century. Technical report, National Science Foundation, Washington, DC. <http://www.nsf.gov/pubs/2005/nsb0540/>.
- McNally, R.; A. Mackenzie; A. Hui; J. Tomomitsu (2012): Understanding the “Intensive” in “Data Intensive Research”: Data Flows in Next Generation Sequencing and Environmental Networked Sensors. In: *IJDC*, 7 (1), 81–94, doi:10.2218/ijdc.v7i1.216.
- Neuroth, H.; S. Strathmann; A. Oßwald; R. Scheffel; J. Klump; J. Ludwig (Hrsg.) (2012): *Langzeitarchivierung von Forschungsdaten – Eine Bestandsaufnahme*.

Boizenburg: Hülsbusch. Online: <http://nestor.sub.uni-goettingen.de/bestandsaufnahme>.

OECD (2006): Recommendation of the Council concerning Access to Research Data from Public Funding. Organisation for Economic Co-operation and Development, Paris.

Pfeiffenberger, H. (2007): Offener Zugang zu wissenschaftlichen Primärdaten. In: *Zeitschrift für Bibliothekswesen und Bibliographie*, 54 (4–5): 207–210. Online: <http://hdl.handle.net/10013/epic.28454.d001>.

Digitale Werkzeuge in den digitalen Geisteswissenschaften

Die Virtuelle Forschungsumgebung TextGrid – Status quo und neue Entwicklungen

von Stefan E. Funk, Ubbo Veentjer und Thorsten Vitt

Einleitung

Als Virtuelle Forschungsumgebung wird TextGrid¹ schon seit einigen Jahren von einer aktiven Community begleitet und als Werkzeug für die tägliche Arbeit genutzt, so zum Beispiel von den Projekten Hybrid-Edition der Fontane-Notizbücher² und Johann Friedrich Blumenbach – online³. Auch Wissenschaftler der Linguistik und der Kunstgeschichte, der klassischen Philologie und der Musikwissenschaft nutzen TextGrid als Virtuelle Forschungsumgebung für ihre Projekte. Eine weitere Zielgruppe von TextGrid sind Entwickler aus dieser Community, die vorhandene Dienste und Werkzeuge von TextGrid für ihre Vorhaben anpassen oder eigene Tools in die Virtuelle Forschungsumgebung integrieren wollen.

Nach der Veröffentlichung von TextGrid 1.0 im Sommer 2011 folgten im Mai 2012 der Betrieb des TextGrid Repositories in Version 2.0 und die zum Download bereitgestellte Version des TextGridLab – mittlerweile in Version 2.0.3. In der gegenwärtigen dritten Förderphase des vom Bundesministerium für Bildung und Forschung (BMBF) geförderten Projektes TextGrid⁴ (2012–2015) wird der nachhaltige Dauerbetrieb der TextGrid-Infrastruktur geplant und umgesetzt. Gemeinsam mit Fachwissenschaftlern und Anwendern sollen das TextGrid Repository mit seinen Diensten sowie das TextGridLab

1 <http://textgrid.de>

2 <http://www.uni-goettingen.de/de/303691.html>

3 <http://blumenbach-online.de/>

4 Förderkennzeichen: 01UG1203A-F

mit seinen Werkzeugen auf eine solide und langfristig nutzbare Basis gestellt werden.⁵

Das TextGrid Laboratory

Das TextGridLab dient als Arbeitsumgebung für Fachwissenschaftler, die digitale Textdaten erstellen und bearbeiten. Es bietet Werkzeuge wie einen XML-Editor zur Erfassung und Auszeichnung von Dokumenten in TEI oder anderen XML-Formaten sowie einen Text-Bild-Linkeditor zur Verknüpfung von Bereichen in digitalisierten Manuskripten mit Textausschnitten in den entsprechenden Transkriptionen. Das TextGrid Repository dient dem TextGridLab unter anderem als Such-, Authentifizierungs- und Autorisierungs- sowie als Storage-Backend.

Tools – Allgemeines

Zum Standard Virtueller Forschungsumgebungen zählen heute Tools zur Verwaltung der Daten und Nutzer. Dementsprechend beinhaltet das TextGridLab eine ganze Reihe solcher Werkzeuge: Dazu gehören ein Navigator zum strukturierten Browsen durch die eigenen Daten, ein Frontend für die Suche sowohl in eigenen wie auch in öffentlichen Daten und eine Benutzerverwaltung, mit der Nutzern Zugriffsrechte auf gemeinsam genutzte Daten zugewiesen werden können.

Eine wichtige Rolle spielt der Metadateneditor: Jedes im TextGridLab erstellte Dokument wird durch einen standardisierten Metadatensatz beschrieben, der nicht nur technische Metadaten wie Größe, Änderungsdatum und Inhaltstyp, sondern auch deskriptive Metadaten enthält.⁶ Dazu gehören unter anderem Titel, beteiligte Personen, beschriebenes Werk und bibliografische Daten einer ggf. digitalisierten Ausgabe. In TextGrid sind ein Dokument und seine Metadaten immer gekoppelt: Lese- und Schreiboperationen liefern beides, verwendet werden die Metadaten unter anderem für Suche und Navigation. Dementsprechend wird im TextGridLab parallel zu geöffneten Dokumenten auch der entsprechende Metadatensatz zur Ansicht und Bearbeitung angeboten. Das Erzeugen von neuen Dokumenten erfordert zumindest eine minimale Angabe von Metadaten vor dem ersten Speichern.

⁵ Vgl. <http://textgrid.de/ueber-textgrid/projekt/>.

⁶ Vgl. http://textgridlab.org/schema/textgrid-metadata_2010.xsd.

Dokumente können über einen Aggregationseditor flexibel gruppiert werden. Damit sind nicht nur die Modellierung einer von Dateisystemen bekannten hierarchischen Verzeichnisstruktur möglich, sondern auch flexiblere Modelle: So können denselben Objekten etwa mehrere Strukturen zugewiesen sein (etwa eine logische und eine physische Struktur). Benutzer können individuelle Sammlungen verteilter Objekte zusammenstellen. Über die Zusammenstellung von Einzelobjekten zu einer Edition können gemeinsame, beispielsweise bibliografische, Metadaten der Objekte zentral am Editionsobjekt verwaltet werden.

Zentrale textwissenschaftliche Tools

XML bearbeiten

Zu den zentralen Tools gehört ein XML-Editor, mit dem Texte etwa im textwissenschaftlichen Auszeichnungsformat TEI (vgl. TEI Consortium 2013) erfasst, bearbeitet und kodiert werden können. Der Editor⁷ bietet entsprechende Funktionalitäten wie Syntax-Highlighting, eine hierarchische Strukturübersicht und Eingabehilfen. Anhand eines assoziierten Schemas können die XML-Dokumente validiert und Fehler angezeigt werden, ebenso ist damit die Auswahl bzw. Autovervollständigung von Tags sowie eine Hilfe zu den jeweils erlaubten XML-Elementen möglich.

Im Rahmen der texttechnologischen Erschließung werden Texte – anders als bei einer reinen Lesefassung – mit recht umfangreichen und komplexen Auszeichnungen versehen. Der TextGrid-XML-Editor bietet deshalb eine Textverarbeitungsprogrammen ähnliche Ansicht, in der XML-Tags wahlweise auf eine symbolische Visualisierung reduziert oder ganz ausgeblendet werden können. Stattdessen wird die XML-Auszeichnung mithilfe eines Subsets von CSS auf Gestaltungsmerkmale wie Schriftarten und -größen, Abstände und Farben abgebildet.

Für diese Funktionalität wurden Teile des visuellen XML-Editors Vex⁸ so angepasst, dass sie als Komponente des XML-Editors eingebunden werden konnten. Dazu gehört unter anderem, dass die vom Benutzer vorgenommene

7 Der XML-Editor basiert auf dem XML-Editor des Eclipse-Webtools-Projekts: <http://www.eclipse.org/webtools/>.

8 <http://wiki.eclipse.org/Vex>

Formatierung des XML-Codes erhalten bleibt, auch wenn der Text in der visuellen Ansicht verändert wird.

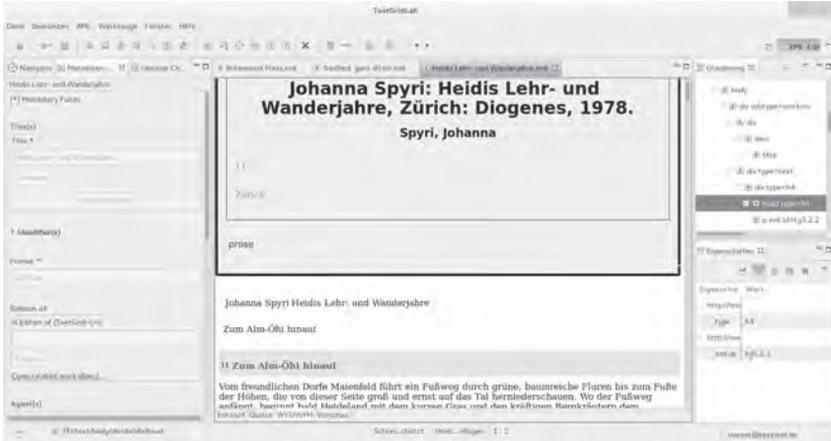


Abb. 1 TextGridLab mit einer WYSIWYM⁹-Ansicht eines TEI-Dokumentes

Neben der visuellen Ansicht, die nur einfache Formatierungen ermöglicht, dafür aber das Bearbeiten des Texts erlaubt, bietet der Editor außerdem einen Preview-Modus: Hier wird das XML-Dokument mit einem vom Benutzer zu spezifizierenden XSLT-Stylesheet in eine HTML-Darstellung transformiert und in einer Browser-basierten Ansicht angezeigt. Dort ist zwar kein direktes Bearbeiten möglich, aber dafür sehr komplexe Formatierungen.¹⁰

Text-Bild-Links

Diverse Editionsprojekte wollen die Digitalisate von Manuskripten etc. mit den entsprechenden XML-Transkriptionen verknüpfen. Für diesen Anwendungsfall wurde der Text-Bild-Link-Editor entwickelt, der dies auf beliebigen Ebenen ermöglicht:

- In einer Ansicht lassen sich im Digitalisat Bereiche bzw. Zonen unterschiedlicher Form und Rotation markieren (z.B. Wörter, Zeilen, Randbemerkungen, etc.) und mit Abschnitten der im XML-Editor geöffneten Transkription verknüpfen.

⁹ WYSIWYM steht für „What You See Is What You Mean“.

¹⁰ Vgl. hierzu den Beitrag zur Hybrid-Edition von Theodor Fontanes Notizbüchern (Radecke/Göbel/Söring 2013) im vorliegenden Band (S. 85 ff.).

- Der Text-Bild-Link-Editor interagiert dabei mit dem oben beschriebenen XML-Editor, indem er die dort vorgenommene Markierung aufnimmt, bei Bedarf Anker Elemente einfügt und die verknüpften Textabschnitte hervorhebt oder anspringt.
- Die Verknüpfungen zwischen Text und Bild an sich werden in einem separaten XML-Dokument auf der Basis von TEI und dem darin einbettbaren Grafikformat SVG¹¹ gespeichert.

Wörterbuchnetz

Zu den ältesten in das TextGridLab integrierten Tools gehört die Anbindung des Trierer Wörterbuchnetzes: Damit ist nicht nur die Suche im Wörterbuchnetz¹² (vgl. Burch/Rapp 2007) mit einer umfangreichen Suchmaske möglich, sondern auch eine direkte Anbindung an den XML-Editor, um beispielsweise das Wort unter dem Cursor automatisch nachzuschlagen.

Weitere, fachspezifische Tools

Neben den oben genannten Kerntools, die zur Standardinstallation des TextGridLab zählen, können eine Reihe weiterer fachspezifischer Werkzeuge zusätzlich im TextGridLab installiert werden. Dazu gehören:

- der Noteneditor MEISE, der die Bearbeitung von Partituren im musikwissenschaftlichen Auszeichnungsformat MEI ermöglicht;
- ein Text-Text-Link-Editor zur Verknüpfung verschiedener Texte untereinander;
- das Tool Digilib¹³ zur Betrachtung und Annotation sehr großer Bilder, ohne dass diese komplett zum Client übertragen werden müssen;
- eine Benutzungsschnittstelle für CollateX¹⁴ zum Kollationieren von Texten;
- ein linguistisches Toolkit mit einem Lemmatizer für das Deutsche und Einbindungen der Korpustools COSMAS II, LEXUS (vgl. Moreira 2011) und ANNEX;
- der SADE-Publisher zur Publikation von in TextGrid bearbeiteten Editionen im Web mithilfe von SADE.

11 <http://www.w3.org/TR/SVG11/>

12 <http://www.woerterbuchnetz.de>

13 <http://digilib.berlios.de/>

14 <http://collatex.net/>

Weitere Informationen zu den einzelnen Tools liefern die TextGrid-Website¹⁵ und das Nutzerhandbuch¹⁶.

Architektur

Dem TextGridLab liegt die aus der Java-Entwicklung stammende modulare und recht weit verbreitete Entwicklungsumgebung Eclipse¹⁷ zugrunde. Das TextGrid-Projekt hat sich für einen Rich Client (d.h. ein lokal zu installierendes Programm, das mit den Servern interagiert) entschieden – anstelle etwa einer Browser-basierten Lösung. Dies liegt vor allem darin begründet, dass es bei Projektbeginn noch deutlich an Software mangelte, um im Browser interaktive Tools zu entwickeln, die es mit Desktop-basierten Programmen in Bezug auf die Bedienung aufnehmen können. Auch heute existiert kein Browser-basierter XML-Editor, dessen Funktionalität mit den verbreiteten Desktop-Lösungen konkurrieren kann.

Als Vorteil von Eclipse ist das reichhaltige Ökosystem an vorhandenen Komponenten hervorzuheben. Als Java-Anwendung ist Eclipse – und damit auch das TextGridLab – plattformübergreifend auf MacOS X, Windows und Linux lauffähig und kann damit eine vielfältige Community ansprechen. Als Open-Source-Projekt steht es samt Quellcode zur Verfügung und kann auch von Dritten weiterentwickelt und modifiziert werden, ohne dass Lizenzkosten anfallen. Es verwendet die jeweils betriebssystemeigenen Widgets und fügt sich damit in das Look and Feel des jeweiligen Betriebssystems ein.

Modularität

Eclipse basiert vollständig auf Plugins – separaten Einheiten, die über definierte Schnittstellen miteinander kommunizieren können und zu Anwendungen zusammengestellt werden. Diese können durch existierende Eclipse-Mechanismen separat installiert und auch aktualisiert werden.

Auf dieser Basis gibt TextGrid mithilfe einer automatischen Update-Funktion Bugfixes und andere Aktualisierungen an Benutzer des TextGrid-

15 <http://textgrid.de/ueber-textgrid/tools-services-ressourcen/>

16 <https://wiki.dariah.eu/wiki/display/TextGrid/User+Manual+2.0>

17 <http://www.eclipse.org/>

Das bearbeitete Schema kommt aus dem TextGridRep, beim Speichern wird direkt dorthin zurückgeschrieben. Auch die Anzeige und Bearbeitung der Metadaten im Metadateneditor parallel zum Editor für den Inhalt ist möglich.

Auf dieselbe Weise lässt sich auch der vom Webtools-Projekt entwickelte CSS-Editor einbinden, der (z.B. als Teil des Web Page Editor) jedoch bei Bedarf nachinstalliert werden muss.

oXygen XML-Editor

Der kommerzielle XML-Editor oXygen¹⁹ ist in der Digital-Humanities-Community weit verbreitet. Im Rahmen von TextGrid ist es nicht möglich (und wohl auch nicht sinnvoll), den eigenen XML-Editor so weit zu entwickeln, dass er die Leistungsfähigkeit von oXygen erreicht. oXygen bietet jedoch eine Eclipse-Plugin-Variante, die sich (eine oXygen-Lizenz vorausgesetzt) ins TextGridLab installieren und dort benutzen lässt.

oXygen überschreibt die entsprechenden Editorzuordnungen, sodass bei installiertem Plugin für unterstützte Inhalte oXygen statt des entsprechenden Standard-TextGridLab-Tools verwendet wird; letztere bleiben via Menü erreichbar.

Integration der Tools

Die Integration dieser Tools wird durch entsprechende Eclipse-Standard-schnittstellen ermöglicht. Im Fall des Dokumentenzugriffs handelt es sich dabei um das Eclipse-Filesystem.

In Eclipse werden Daten in einem Arbeitsbereich verwaltet, in dem es Projekte und darunter ein hierarchisches System von Ordnern und Dateien gibt. Der Zugriff darauf erfolgt über eine Eclipse-eigene Abstraktionsschicht, die z.B. von Editoren genutzt wird.

TextGrid arbeitet mit Objekten, die über eine (opake) URI – die TextGrid-URI – identifiziert und durch Metadaten beschrieben werden. Der Zugriff erfolgt über eine allgemeine, aber auf die Spezifika von TextGrid abgestimmte (vulgo proprietäre) Webservice-Schnittstelle.

¹⁹ <http://www.oxygenxml.com/>

Greift ein Benutzer z.B. über den Navigator auf ein TextGrid-Objekt zu, spricht ein entsprechender Bestandteil des TextGridLab die TextGrid-Schnittstellen an und erzeugt ein virtuelles Dateiojekt im Eclipse-Arbeitsbereich, über das Eclipse-Editoren dann mit den Daten arbeiten können, ohne die TextGrid-Schnittstellen zu kennen. Dabei werden auch gleich die Metadaten geladen bzw. gesichert, die im Metadateneditor parallel zum aktuellen Editor dargeboten werden.

Kollaboratives Arbeiten mit ECF

Jedoch lassen sich nicht nur Editoren für besondere Dokumenttypen einbinden – auch andere, externe Tools sind integrierbar. Ein Beispiel liefert das Eclipse Communication Framework (ECF)²⁰, das kollaboratives Arbeiten in Echtzeit ermöglicht.

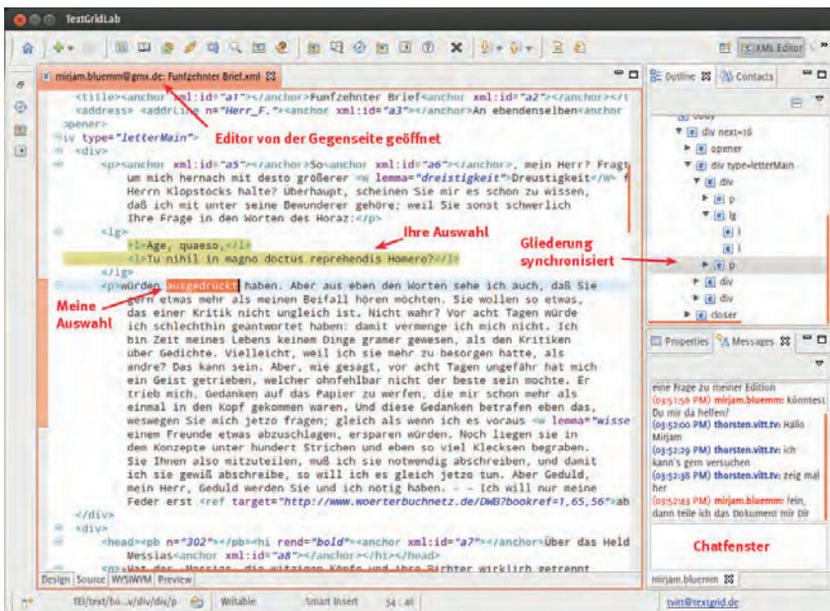


Abb. 3 Gemeinsam von zwei Personen im XML-Editor bearbeiteter Text. Bearbeitungen erscheinen unmittelbar bei beiden.

²⁰ <http://www.eclipse.org/ecf/>

Nach der Installation des ECF ins TextGridLab kann der Benutzer sich bei einem Chat-Anbieter (z.B. einem XMPP-Anbieter wie Jabber.org, Google Talk, GMX oder Rechenzentren) anmelden. So ist mit den entsprechenden Kontakten ein Chat direkt im TextGridLab-Fenster möglich. Viel interessanter ist jedoch die DocShare-Funktion: Via Kontextmenü kann der Benutzer beliebige Editoren an seinen Kontakt senden. Der Editor erscheint dann auch im TextGridLab-Fenster seines Kontakts, beide Parteien können gleichzeitig am selben Dokument arbeiten und live die Änderungen des Kollegen verfolgen.

Da das ECF explizit auf Erweiterbarkeit durch andere Tools ausgelegt ist, sind mit ein wenig Programmierarbeit Szenarien denkbar, die die kollaborativen Fähigkeiten stärker ins Lab integrieren. So könnte z.B. mit einem selbst betriebenen XMPP-Server jedem TextGrid-Benutzer ein XMPP-Account zugeordnet werden und die Anmeldung am Server erfolgte dann automatisch mit der Anmeldung oder über einen „Online“-Knopf in der Toolbar. Entsprechend könnte beispielsweise für Projekte ein Gruppenchat angeboten werden, in dem alle zurzeit aktiven Projektmitglieder miteinander diskutieren. Ebenso könnte bei Bearbeitungskonflikten (A möchte ein Objekt bearbeiten, das B bereits geöffnet hat) das weitere Vorgehen über ein Chatfenster diskutiert und das Objekt z.B. im oben beschriebenen gemeinsamen Modus geöffnet werden; über die Anbindung z.B. eines TextGrid-weiten Benutzerchats wäre die Möglichkeit eines Live-Supports evtl. auch der Benutzer untereinander gegeben.

Das TextGrid Repository

TextGrid bietet mit dem TextGridRep ein Repository für geisteswissenschaftliche Forschungsdaten und fungiert so als fachwissenschaftliches Langzeitarchiv für die textbasierten Geisteswissenschaften. Die über das TextGridRep angebotenen Dienste werden unter anderem als Grundlage für die digitale Arbeitsumgebung TextGridLab genutzt.

Neben dem Zugang zum TextGridRep im World Wide Web,²¹ der einen Zugriff auf alle im TextGrid Repository archivierten Daten erlaubt, wird es perspektivisch an die Anforderungen der Projekte angepasste Portale geben,

21 Vgl. <http://textgridrep.de/>.

um prominenten und gezielten Zugriff auf deren Forschungsdaten zu ermöglichen.

Das TextGrid Repository bietet derzeit Teile der Digitalen Bibliothek an, einer umfangreichen digitalen Textsammlung vom Anfang des Buchdrucks bis zu den ersten Jahrzehnten des 20. Jahrhunderts. Diese Textsammlung hat TextGrid als Online-Bibliothek von Zeno.org²² erworben. Die Texte werden von TextGrid nach TEI aufbereitet und mit zusätzlichen Metadaten angereichert und stehen so nicht nur zur Lektüre, sondern auch zur Weiterverarbeitung in Editionen und Korpora zur Verfügung.²³ Sie können unter der Creative-Commons-Lizenz-„by“-Version 3.0²⁴ kostenfrei nachgenutzt werden.

Speichern und Manipulieren von Daten

Zentral für ein digitales Repository sind sicherlich Dienste zum Speichern oder generell zur Manipulation von Daten und ihrem komfortablen und schnellen Wiederauffinden. Ersteres leistet TextGrid durch den Speicherdienst TG-crud²⁵ – ein Dienst, der unter anderem die CRUD-Operationen Create, Retrieve, Update und Delete unterstützt, also das Erstellen, Lesen, Aktualisieren und Löschen von Daten.

Suchen und Finden

Zum schnellen Suchen und Finden von Dokumenten auch in großen Datenmengen bietet TextGrid den Suchdienst TG-search an, der alle relevanten Daten des Repositories indiziert und eine komfortable Metadaten- und Volltextsuche zur Verfügung stellt. Beide Dienste verfügen über dokumentierte Schnittstellen nach außen.²⁶ Sie werden so vom TextGridLab und auch der Website des TextGrid-Repositories angesprochen. Eine Nutzung dieser Schnittstellen für die Suche und den Zugriff auf Daten des TextGridRep für externe Dienste ist ebenfalls möglich.

22 <http://zeno.org>

23 Vgl. <http://textgrid.de/ueber-textgrid/digitale-bibliothek/>.

24 <http://creativecommons.org/licenses/by/3.0/de/>

25 <https://wiki.dariah.eu/wiki/display/TextGrid/TG-crud>

26 Vgl. <https://wiki.dariah.eu/wiki/display/TextGrid/TG-crud>, <https://wiki.dariah.eu/wiki/display/TextGrid/TG-search>.

Rechtmanagement

Um Personen²⁷ oder Gruppen von Personen einen gezielten Zugriff auf Dokumente zu gestatten, ist ein Rechtmanagement erforderlich. Ausgenommen sind hier die öffentlichen Dokumente des TextGrid-Repositories; diese sind, einmal publiziert, für alle verfügbar. Das Rechtmanagement stellt die detaillierte und sichere Authentifizierung und Autorisierung auch auf nicht öffentliche Daten sicher. Der Authentifizierungs- und Autorisierungsdienst TG-auth*²⁸ ermöglicht dies mithilfe der Single-Sign-On-Lösung Shibboleth²⁹ und dem rollenbasierten Zugriffssystem OpenRBAC³⁰. Arbeiten mit Dokumenten in TextGrid sind diesbezüglich abgesichert, indem verschiedene Rollen innerhalb eines Projektes verschiedene Rechte an einzelnen Dokumenten des Projektes innehaben: Eine Person mit der Rolle „Bearbeiter“ darf unter anderem Dokumente erstellen und bearbeiten, ein „Beobachter“ Dokumente nur lesen, die Person mit der Rolle „Projektleiter“ außerdem Rollen zuteilen (delegieren) und nur Nutzer mit der Rolle „Authority to delete“ dürfen Dokumente löschen.

Importieren in das TextGridLab – der dynamische Speicherbereich des TextGridRep

TextGridLab-Import-Modul

Das im TextGridLab bereitgestellte Import-Modul ermöglicht den Import lokaler Dateien und auch ganzer lokaler Verzeichnisstrukturen in den dynamischen Bereich des TextGrid-Repositories (siehe Ingest 1, Abb. 4). Der dynamische Bereich ist derjenige, in dem Dokumente abgelegt werden, die vom TextGrid-Rechtmanagement verwaltet werden und somit nur für angemeldete und freigeschaltete Personen sichtbar sind. Dieser Speicherbereich wird für eine kontinuierliche und kollektive Arbeit an nicht öffentlichen Dokumenten benutzt. Die über das Import-Modul des TextGridLab eingespiel-

27 Mit Personen sind hier diejenigen gemeint, die entweder einen TextGrid-Account beantragt und erhalten haben oder sich per Shibboleth bei TextGrid anmelden.

28 Das „*“ in TG-auth* steht hier zum einen für „N“ – Authentification, und zum anderen für „Z“ – Authorization.

29 <http://shibboleth.net/>

30 <http://www.openrbac.de/>

ten Dokumente werden alle über den Speicherdienst TG-crud in diesen dynamischen Speicherbereich geschrieben – sowie in den Suchindex (siehe Suchindex 1, Abb. 4), für den die XML-Datenbank eXist³¹ und die RDF-Datenbank Sesame³² eingesetzt werden. Schließlich werden alle Dokumente beim Rechtemanagement registriert und sind so vor unerlaubtem Zugriff geschützt. Beim Import von Daten in das TextGridLab werden in den Dokumenten enthaltene Referenzen auf andere zu importierende Dokumente von lokalen Dateipfaden nach (während des Einspielvorgangs automatisch generierten) TextGrid-URIs umgeschrieben, sodass diese Referenzen innerhalb des TextGridRep ihre Gültigkeit nicht verlieren.

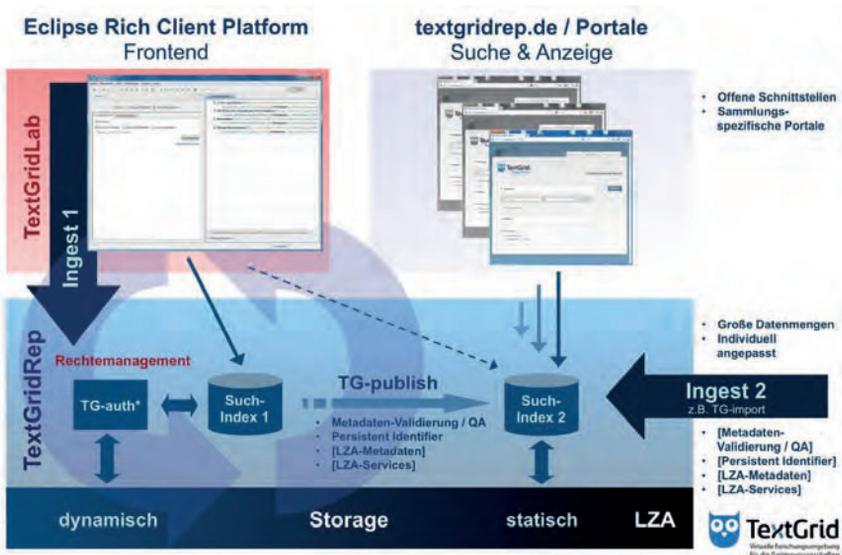


Abb. 4 Die TextGrid-Repository-Architektur

Importieren mit TG-import

Eine weitere Möglichkeit, Daten in den dynamischen Bereich des TextGrid-Rep zu importieren – um mit diesen Dokumenten dann im TextGridLab zu arbeiten –, bietet die Software-Bibliothek TG-import (s. Ingest 2, Abb. 4).³³

31 <http://exist-db.org/>

32 <http://www.openrdf.org/>

33 Vgl. <https://wiki.dariah.eu/wiki/x/1oBn>.

Diese Methode dient zuvorderst dem direkten Import – der Publikation – auch großer Datenmengen in das öffentliche TextGrid Repository, mithin in den statischen Speicherbereich und Suchindex 2. Durch Konfiguration ist es jedoch möglich, diesen Ingest 2 auch für den Datenimport in den dynamischen Speicherbereich des TextGridRep und den Suchindex 1 zu nutzen. So können auch Daten im TextGridLab verfügbar gemacht werden, die beispielsweise individuell vorbearbeitet werden. TG-import erweitert die in den Projekten kopal³⁴ und DP4lib³⁵ entwickelte Java-Bibliothek koLibRI³⁶ (kopal Library of Retrieval and Ingest) mit diversen Modulen und Diensten für die Nutzung mit TextGrid.

Ebenso wie im TextGridLab-Import-Modul besteht die Möglichkeit, ganze Verzeichnisstrukturen einzuspielen, wobei die Struktur des Quellordners übernommen wird. Für fortgeschrittene Nutzer bzw. Entwickler mit TextGrid-Objekt-, XML- und XSLT-Kenntnissen kann auch ein Import von fertig vorbereiteten Datenobjekten direkt durchgeführt werden, wobei dann die TextGrid-Metadaten, Referenzen, TextGrid-URIs und weitere benötigte und gewünschte Datenstrukturen selbst erzeugt werden können bzw. vor dem Importieren vorliegen müssen; dies ermöglicht einen sehr flexiblen Import.

Darüber hinaus ist ein Einspielen von DFG-Viewer³⁷-METS³⁸-Dateien möglich, wobei die dort enthaltenen Metadaten und referenzierten Dokumente (zum Beispiel Scans von Buchseiten und/oder Handschriften) heruntergeladen und in das TextGridRep kopiert werden. In einer METS-Datei können generell Struktur- und Metadaten eines digitalen Objekts strukturiert beschrieben werden, in einer DFG-Viewer-METS-Datei sind diese Strukturen nach einer bestimmten Spezifikation abgelegt – dem zvd/DFG-Viewer METS-Profil.³⁹ Dies beinhaltet eine festgelegte Struktur sowie bestimmte Metadaten und auch Referenzen auf die Digitalisate selbst. So können die Digitalisate mitsamt ihrer Struktur und deren dazugehörigen Metadaten mittels des DFG-Viewers anschaulich dargestellt werden. Beim Import der

34 <http://kopal.langzeitarchivierung.de>

35 <http://dp4lib.langzeitarchivierung.de/>

36 http://dp4lib.langzeitarchivierung.de/index_koLibRI.php.de

37 <http://dfg-viewer.de/>

38 Metadata Encoding and Transmission Standard (METS); vgl. <http://www.loc.gov/standards/mets/>.

39 Vgl. <http://dfg-viewer.de/profil-der-metadaten/>.

Daten in das TextGridRep werden die Referenzen so umgeschrieben, dass die dort enthaltene METS-Datei auf die importierten Digitalisate desselben verweisen. Außerdem werden die in der METS-Datei enthaltene logische und physikalische Struktur sowie alle gewünschten Metadaten, konfigurierbar über XSLT-Stylesheets, übernommen.

Publizieren mit TextGrid – der statische Speicherbereich des TextGridRep

Publizieren aus dem TextGridLab

Das TextGridLab bietet die Möglichkeit, zum Beispiel kritische Editionen kollaborativ und an verteilten Standorten zu erstellen. Ist diese Arbeit abgeschlossen, kann sie vom TextGridLab aus veröffentlicht werden. Eine solche Publikation in TextGrid umfasst die Metadaten-Kontrolle aller enthaltenen Dokumente und bietet eine zitierfähige und dauerhafte Speicherung der Daten. Für diesen Vorgang ist der Dienst TG-publish zuständig (s. TG-publish, Abb. 4).

Ebenfalls als Erweiterung der Bibliothek koLibRI implementiert, wird TG-publish vom TextGridLab aus aufgerufen und alle erforderlichen Schritte werden in der Publish-GUI⁴⁰ des TextGridLab Schritt für Schritt durchgeführt (s. Abb. 5).

Korrekte und vollständige Metadaten vorausgesetzt, wird für jedes Dokument ein persistenter Identifikator (Persistent Identifier – PID) erzeugt. TextGrid nutzt hierfür Handles als PIDs, die im Rahmen von EPIC⁴¹ von der GWDG⁴² vergeben werden. Schließlich werden alle Dokumente von Suchindex 1 in Suchindex 2 kopiert, in den statischen Speicherbereich geschoben und die Zugriffsrechte auf „öffentlich lesbar“ gesetzt. Nun sind die Daten im TextGridRep publiziert und auch auf dessen Internetseite nachgewiesen. Sie können dort ebenso wie auch im Navigator des TextGridLab per Browsing und per Suche recherchiert werden.

40 <https://wiki.dariah.eu/wiki/display/TextGrid/Publish>

41 European Persistent Identifier Consortium; vgl. <http://www.pidconsortium.eu/>.

42 Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen.; vgl. <http://gwdg.de>

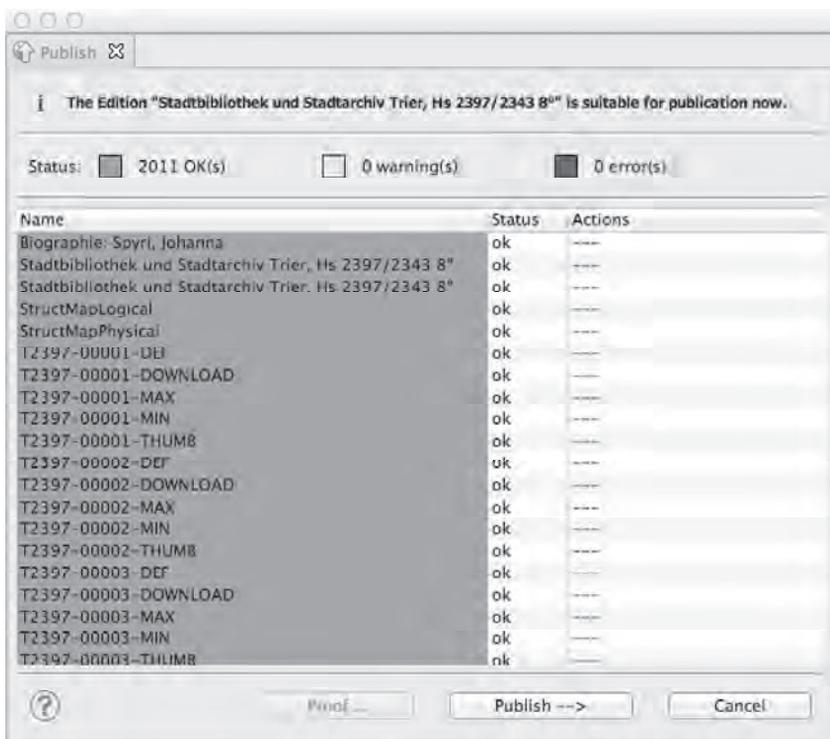


Abb. 5 Die TG-publish-GUI des TextGridLab – nach einem erfolgreichen Prüfen der Voraussetzungen für den Publikationsvorgang

Direktes Publizieren in das TextGridRep

Ebenso wie in Abschnitt „Importieren mit TG-import“ beschrieben, können mit TG-import (s. Ingest 2, Abb. 4) Daten auch direkt in den statischen Bereich des TextGrid-Repositories importiert werden. Dies kommt prinzipiell einer Publikation gleich, die aus dem TextGridLab heraus ausgeführt wird, nur dass zum einen der Import- bzw. Publikations-Vorgang flexibler gestaltet werden kann (durch den modularen Aufbau von TG-import bzw. koLibRI) und zum anderen keine Daten verschoben oder kopiert werden, sondern direkt in den Suchindex 2 und den statischen Speicherbereich des TextGridRep geschrieben werden. Per TG-import importierte Daten werden zunächst in eine sogenannte Sandbox publiziert – einen Bereich des TextGridRep, der auch öffentlich einsehbar ist, dessen Dokumente jedoch noch nicht den finalen Status „publiziert“ innehaben. Hier können sämtliche Dokumente noch-

mals auf Korrektheit und evtl. fehlerhaft durchgeführte Prozesse bei der Publikation hin überprüft werden – zum Beispiel, ob alle internen TextGrid-Identifizierer (die TextGrid-URIs) korrekt in PIDs umgeschrieben wurden. Sind die Daten fehlerfrei, können sie mithilfe von TG-import final publiziert oder andernfalls aus der Sandbox gelöscht werden.

Publizieren eigener digitaler Editionen mit SADE

Das TextGridRep bietet einen generischen Zugang zu den publizierten Daten, mit Navigations- und Suchmöglichkeiten, die möglichst für alle Projekte nutzbar sind. Zahlreiche Forschungsvorhaben möchten jedoch eigene Darstellungs- und Suchmöglichkeiten für ihre Daten im Web schaffen, individuelle Oberflächen gestalten oder auch eigene Werkzeuge für die Interaktion mit den Daten bereitstellen. Hierfür bietet TextGrid die Publikationsmöglichkeit in eine eigene SADE⁴³-Installation an. Dabei handelt es sich um eine Art Baukasten zur Publikation digitaler Editionen, der Module für die Suche in und die Darstellung von TEI-Volltexten liefert. Projekte können weitere Module wie beispielsweise eine Raum-Zeit-Präsentation entwickeln und auch mit der Community teilen. Da SADE auf der XML-Datenbank eXist⁴⁴ basiert, können für die Entwicklung eigener Module XML-Technologien wie XQuery und XSLT eingesetzt werden. Wenn Daten, die auch im TextGrid Repository veröffentlicht wurden, via SADE publiziert werden, sichert der vergebene persistente Identifikator (PID) die Zitierfähigkeit durch langfristige Verfügbarkeit der Originaldaten im TextGrid Repository. Entwickelt wird SADE an der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW); inzwischen hat sich eine größere Community gebildet, die sich aktiv an der Weiterentwicklung beteiligt.

Dienste für die Langzeitarchivierung

Als fachwissenschaftliches Langzeitarchiv bietet das TextGrid Repository zurzeit eine Metadaten-Validierung beim Import aus dem TextGridLab, eine Bitstream-Preservation der archivierten Daten sowie eine dauerhafte Zitierfähigkeit der Dokumente per Persistent Identifier, für deren Pflege in der gegenwärtigen Förderphase die SUB Göttingen verantwortlich zeichnet. Langfristig ist eine Integration des TextGrid-Repositories in das Angebot der

43 Scalable Architecture for Digital Editions; vgl. <http://www.bbaw.de/telota/software/sade/>.

44 <http://exist-db.org/>

Infrastruktur DARIAH-DE vorgesehen, um die nachhaltige Zugänglichkeit der Forschungsdaten sowohl auf technischer wie auf struktureller und auch fachwissenschaftlicher Ebene zu gewährleisten. Weitere Dienste für die Langzeitarchivierung werden in Zukunft entwickelt werden und betreffen beispielsweise sowohl die Extraktion von technischen Metadaten und die Formatvalidierung während des Imports wie auch Module für die Formatmigration von Dokumenten. Für eine umfassende Data Curation ist auch eine detaillierte Planung der Langzeiterhaltung – das Preservation Planning – unerlässlich.

Das TextGrid Repository im World Wide Web – Werkzeuge und Dienste

Mit TextGrid veröffentlichte Dokumente können auf der Internetseite des TextGrid-Repositories – textgridrep.de – eingesehen werden. Dort besteht die Möglichkeit, Dokumente mit einer einfachen sowie einer erweiterten Suche zu recherchieren. Eine facettrierte Suche ist für eine zukünftige Version der Website vorgesehen. Über die Seite „Repository Content“ gibt es die Möglichkeit, durch alle veröffentlichten Editionen zu navigieren. Veröffentlichte TEI-Texte und Bilder können direkt in der Repository-Ansicht betrachtet werden. Einzeldokumente können als TEI-Dokument oder als E-Book im EPUB-Format heruntergeladen werden. Ganze Textsammlungen sind außerdem als TEI-Corpus zum Download verfügbar.

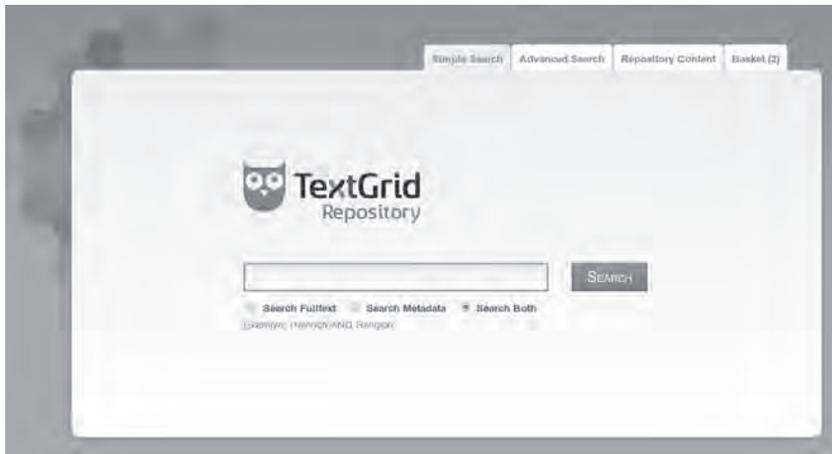


Abb. 6 Der Einstieg in die Website textgridrep.de

Tools⁴⁶ zu analysieren. Die Voyant-Tools bieten eine webbasierte Arbeitsoberfläche, die verschiedene Werkzeuge zur statistischen Textanalyse auf Einzeldokumente oder Korpora bietet, darunter Ansichten von Wordclouds, Termhäufigkeiten, Graphen und vieles mehr.

Für die Anbindung der TextGridRep-Website an Voyant wurde von DARIAH-DE ein Warenkorb in die Website integriert, in dem gewünschte Texte gesammelt werden können. So wird auch aus der TextGridRep-Website ein Labor für die Analyse von publizierten Texten im Browser.

Synopsentool – CollateX

The screenshot shows the 'TextGrid Collate Result' interface. At the top, there are navigation tabs: 'Simple Search', 'Advanced Search', 'Repository Content', 'Basket (3)', and 'Collate Result'. The main content area displays a comparison of three versions of a text passage from 'Kapitel 41 des Blumenbachschen „Handbuchs der Naturgeschichte“'. The text is presented in a table with three columns, each representing a different version of the text. The text is highlighted in yellow to show differences between the versions.

00023_1782_2_Saeugetiere (textgrid:208e8.0)	00027_1799_8_Saeugetiere (textgrid:208e9.0)	00033_1825_11_Saeugetiere (textgrid:208eb.0)
§. 41. Die Säugethiere haben	§. 41. Die Säugethiere haben	§. 41. Die Säugethiere haben
zwar		
das warme rolhe Blut mit den Vögeln gemein;	das warme rolhe Blut mit den Vögeln gemein;	das warme rolhe Blut mit den Vögeln gemein;
doch zeichnen	aber	aber
sie	sie	sie
sich schon dadurch von ihnen aus, daß sie keine Eyer legen, sondern	gebären	gebären
lebendige Junge	lebendige Junge:	lebendige Junge:
gebähren:	und	und
ihr Hauptcharakter	ihr Hauptcharakter,	ihr Hauptcharakter,
aber,		
der sie von allen übrigen Thieren unterscheidet, und von dem auch die Benennung der ganzen Classe entlehnt ist, sind die Brüste, wodurch die	der sie von allen übrigen Thieren unterscheidet, und von dem auch die Benennung der ganzen Classe entlehnt ist, sind die Brüste, wodurch die	der sie von allen übrigen Thieren unterscheidet, und von dem auch die Benennung der ganzen Classe entlehnt ist, sind die Brüste, wodurch die
Weibgen	Weibchen	Weibchen
ihre Junge mit Milch ernähren.	ihre Junge mit Milch ernähren.	ihre Junge mit Milch ernähren.

Abb. 8 Die Ausgabe des Synopsentools (Prototyp) mit dem Vergleich dreier Versionen des Beginns von Kapitel 41 des Blumenbachschen „Handbuchs der Naturgeschichte“.

Auf derselben Grundlage soll auch ein Synopsentool in die Website integriert werden, das im Rahmen des „Blumenbach – online“-Projektes entwickelt wird. Das Synopsentool basiert auf CollateX⁴⁷ und ermöglicht einen

⁴⁶ <http://voyant-tools.org/>

⁴⁷ <http://collatex.net/>

übersichtlichen Vergleich mehrerer Versionen eines Textes (s. Abb. 8). Hierzu können verschiedene Texte, die im TextGridRep veröffentlicht wurden, im oben bereits erwähnten Warenkorb gesammelt und kollationiert werden.⁴⁸

Shared Canvas – Annotationen auf Repository-Daten

Im Rahmen des von der Andrew W. Mellon Foundation⁴⁹ geförderten Projektes SharedCanvas⁵⁰ sollen im Zeitraum von Mai bis November 2013 Daten des TextGrid-Repositories für SharedCanvas-Tools nutzbar gemacht werden. SharedCanvas ermöglicht die Annotation von Bilddokumenten in SharedCanvas-kompatiblen Repositorien. Dabei handelt es sich um sogenannte Stand-off-Annotationen, bei denen die Originaldokumente unmodifiziert bleiben. Hierzu integriert SharedCanvas den OpenAnnotation⁵¹-Standard. So können ein Bilddokument aus dem TextGrid Repository – beispielsweise ein Scan eines mittelalterlichen Textes – in SharedCanvas-Tools neben einem Bilddokument aus einem Repository der Stanford University betrachtet und Anmerkungen zu beiden Dokumenten in einem dritten Annotationsdokument festgehalten werden. Für die Darstellung der Bilddokumente definiert SharedCanvas einen eigenen Bilderservice⁵², der es ermöglicht, Bilder aus dem Repository in verschiedenen Größen, Qualitätsstufen, Formaten oder Rotationswinkeln zu adressieren. Von TextGrid soll hierzu das bereits integrierte Werkzeug DigiLib angepasst werden.

Diese Integration bietet für die SharedCanvas-Community den Vorteil, dass Bildobjekte des TextGrid-Repositories für SharedCanvas-Werkzeuge verfügbar sind. TextGrid-Nutzer profitieren durch die Annotationswerkzeuge, mit denen Bilder der beteiligten Repositorien gemeinsam annotiert werden können. Zu den Unterstützern des gemeinsamen Standards gehören unter anderem die Stanford University, die British Library, die Bodleian Libraries (Oxford University) und die Bibliothèque Nationale de France.⁵³

48 Vgl. hierzu den Beitrag zur Edition Blumenbach Online (Kerzel/Reich/Weber 2013) im vorliegenden Band (S. 107 ff.).

49 <http://www.mellon.org/>

50 <http://www.shared-canvas.org/>

51 <http://www.openannotation.org/>

52 IIIF. International Image Interoperability Framework; vgl. <http://www.sul.stanford.edu/iiif/image-api/>.

53 Vgl. <http://iiif.io/community.html>.

Ausblick

TextGrid befindet sich nunmehr in der dritten Projektphase: Während der technische Fokus der ersten Phase auf Konzeption und prototypischem Aufbau der Virtuellen Forschungsumgebung lag und im zweiten Abschnitt TextGrid in den Produktivbetrieb überführt, modularisiert und für weitere Fachdisziplinen spezifiziert wurde, zeigt eine stetig wachsende Anzahl an Forschungsvorhaben heute dezidiertes Interesse, TextGrid für eigene Projekte und Aufgaben praktisch zu nutzen und eigene Infrastruktur anzubinden oder befindet sich bereits in der konkreten Umsetzung. Dies gilt es im Blick auf künftige Entwicklungen ebenso zu berücksichtigen wie die Anbindung an generischere Infrastrukturprojekte wie DARIAH-DE.

Im Bereich der Repository-Infrastruktur liegt der Fokus vor allem auf der Verbesserung von Performance und Stabilität, um einen nachhaltigen Betrieb zu gewährleisten. So sind eine differenzierte Qualitätssicherung ebenso wie die Parallelisierung der TextGridRep-Dienste für eine erhöhte Ausfallsicherheit des TextGrid-Repositories vorgesehen (vgl. Berber et al. 2013.). Eine weitere Aufgabe wird es sein, den Publikations-Mechanismus zu verfeinern und weiter auszubauen.

Die TextGridRep-Website wird zunehmend mit Tools zu Aufbereitung und Analyse ausgestattet (Voyant, Korpuszusammenstellung, Export in Formate wie TEIcorpus und EPUB, Synopsentool – weitere werden folgen). Dies legt auch hier eine stärkere Modularisierung nahe. Zudem bietet sich an, diese Werkzeuge auch für noch nicht veröffentlichte Daten zu Verfügung zu stellen.

Es erscheint deshalb sinnvoll, eine gemeinsame Weboberfläche für den statischen und dynamischen Bereich zu entwickeln, die vom Blättern, Suchen, Lesen und Auswerten der im TextGridRep veröffentlichten Inhalte ausgeht, diese nach erfolgter Anmeldung jedoch um den Zugriff auf eigene, geschützte Inhalte erweitert. Weitere Funktionen können dieses Angebot sinnvoll um interaktive Angebote ergänzen: Dazu zählen etwa Annotationsfunktionen⁵⁴, Korpusaufbereitung und -zusammenstellung, aber auch einfachere Bearbeitungsfunktionen. Das TextGridLab als Rich Client bietet so eine modulare Plattform für Aufgaben, die komplexe Bearbeitungen umfassen und interaktive Spezialtools etwa aus dem Eclipse-Ökosystem erfordern.

54 Vgl. SharedCanvas; aber auch der Text-Text-Link-Editor im TextGridLab, derzeit ein Hybrid, ist auf den Einsatz als Webtool vorbereitet.

In den digitalen Geisteswissenschaften hat sich in den letzten Jahren viel bewegt und TextGrid hat wesentlich dazu beigetragen, Fachwissenschaftlern diesen neuen Zugang – oder überhaupt einen ersten Zugang – zu den digitalen Medien zu ermöglichen; sei es im Zusammenhang mit zum Beispiel großen Textmengen oder auch der Annotation von Texten und/oder Bildern. TextGrid bietet heute mit seiner Virtuellen Forschungsumgebung – dem TextGridLab, dem TextGridRep und allen dazugehörigen Diensten – eine umfassende Grundlage an Werkzeugen und Diensten, die für neue Forschungsfelder und -fragen genutzt werden können und in der Praxis bereits umgesetzt werden. Die Beispiele in diesem Beitrag zeigen deutlich, dass es für die digitalen Geisteswissenschaften noch viele Möglichkeiten gibt, die Dinge weiter zu entwickeln. Auch die Kooperation und die Nutzung von Synergien, beispielsweise mit dem Projekt DARIAH-DE und dessen Ziel, „die mit digitalen Ressourcen und Methoden arbeitenden Geistes- und Kulturwissenschaftler in Forschung und Lehre zu unterstützen“ (DARIAH-DE 2012: 3), zeigt, dass noch weitere Partner im Boot sitzen und kräftig rudern. TextGrid wird mit dabei sein.

Literaturverzeichnis

- Albritton, Ben; Peter James, Neil Jefferies, Christopher Jesudurai, Sean Martin, Roger Mathisen, Petter Rønningsen (2012): IIIF. International Image Interoperability Framework: Image API 1.0, 10. August 2012. <http://www-sul.stanford.edu/iiif/image-api/>.
- Berber, Fatih; Stefan E. Funk, Ubbo Veentjer, Martin Haase, Philipp Wieder, Peter Gietz (2013): Architekturskizze inklusive Identifikation der Lücken und der Optimierungsmöglichkeiten (R 4.5.1), 22. Februar 2013. http://www.textgrid.de/file-admin/user_upload/TextGrid_R451_Architekturskizze.pdf.
- Burch, Thomas; Andrea Rapp (2007): Das Wörterbuch-Netz: Verfahren – Methoden – Perspektiven. In: *Historisches Forum* 10 (1). Online: http://edoc.hu-berlin.de/histfor/10_I/PHP/Woerterbuecher_2007-10-I.php#007001.
- DARIAH-DE (Hrsg.) (2012): DARIAH-DE. Aufbau von Forschungsinfrastrukturen für die eHumanities. Göttingen, Online: <http://dx.doi.org/10.3249/webdoc-3589>.
- Kerzel, Martina/Reich, Mike/Weber, Heiko (2013): Die Edition „Johann Friedrich Blumenbach – online“ der Akademie der Wissenschaften zu Göttingen. In: Neuroth, Heike; Lossau, Norbert; Rapp, Andrea (Hrsg.): *Evolution der Informationsinfrastruktur. Forschung und Entwicklung als Kooperation von Bibliothek und Fachwissenschaft*. Glückstadt: Hülsbusch, S. 107–136 [in diesem Band].

- Moreira, André (2011): LEXUS on the TextGrid infrastructure – exploring new potentialities. In: *Language Archiving Technology – News*, Dezember 2011. <http://tla.mpi.nl/tla-news/lexus-on-the-textgrid-infrastructure-exploring-new-potentialities/>.
- Radecke, Gabriele; Göbel, Mathias; Söring, Sibylle (2013): Theodor Fontanes Notizbücher. In: Neuroth, Heike; Lossau, Norbert; Rapp, Andrea (Hrsg.): *Evolution der Informationsinfrastruktur. Forschung und Entwicklung als Kooperation von Bibliothek und Fachwissenschaft*. Glückstadt: Hülsbusch, S. 85–105 [in diesem Band].
- TEI Consortium (Hrsg.) (2013): TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 2.3.0, 17. Januar 2013. <http://www.tei-c.org/Guidelines/P5/>.

Technische Infrastrukturen, Dienste und Services für die digitalen Geisteswissenschaften

von Peter Wittenburg und Peter Gietz

Einleitung

Wir leben in einer Phase, in der uns in vielen Fällen nur *komplexe Systeme* weiterhelfen, um die Aktionen der Menschen auf die Natur und Umwelt, auf die Gesellschaften und auf das Individuum ausreichend tiefgründig zu verstehen. Das ESFRI-Roadmap-Dokument (ESFRI 2006) spricht in diesem Zusammenhang vom Folgenden:

- Wenn Menschen eine neue Initiative starten wollen, müssen sie vorher-sagen können, welche Folgen eine solche Initiative hat.
- Was immer wir auch unternehmen, es wird erhebliche Konsequenzen auf die Natur, den Geist und die Gesellschaften in verschiedenen Dimensio-nen haben. Daher bedarf es komplexer Modelle, um die Interdependen-zen gut zu verstehen.

Das Dokument schließt bewusst die Verfassung des menschlichen Geistes und der Gesellschaft mit in den Kanon der großen Probleme ein und dies ist die Domäne der Geistes-, der Kunst- und der Sozialwissenschaften. Es wird zu Recht sehr viel über die Fragen diskutiert, wie wir die Stabilität des Kli-mas, der Energieversorgung, der Gesundheit etc. in der Zukunft garantieren können. Ebenso wichtig sind aber die Fragen nach dem Erhalt der Stabilität des menschlichen Geistes und der Gesellschaften angesichts der enormen, weltweiten Migrationsbewegungen, die zu einer Erosion gewachsener Kultu-ren und Sprachen führen, und angesichts einer immer stärker werdenden technologischen Innovationsrate, die viele Menschen vor immense Probleme stellt.

Die Geistes- und Sozialwissenschaften (GSW) stehen mithin vor ähnlich *großen Herausforderungen*, die es zu bewältigen gilt, wie z.B. die Naturwis-senschaften. Mithin sind auch die Methoden, die erforderlich sind, um diese Herausforderungen aufzugreifen, von ähnlicher Art. Wissenschaftler der GSW müssen einen einfachen und flexiblen Zugriff auf die verschiedensten digitalen Daten und Werkzeuge haben – und das disziplin- und länderüber-

greifend. Sie müssen sich *neue Paradigmen* der Herangehensweise erarbeiten, die es – wie auch z.B. in der Naturwissenschaft die „data-intensive science“, also das Operieren auf großen Datenmengen unterschiedlicher Herkunft mittels statistischer/stochastischer Algorithmen – erlauben, komplexe Muster zu finden, die dem Wissenschaftler anders verborgen geblieben wären.

Vor einigen Jahren wurde der Begriff der *E-Science* bewusst vage geprägt, um diese neuen Herausforderungen und auch Möglichkeiten anzudeuten. Zu Recht wurde analog hierzu der Begriff *E-Humanities* geprägt, wobei wir an dieser Stelle nicht auf die Nuancen eingehen wollen, die die Begriffe „E-Humanities“¹, „Digital Humanities“ und „Computational Humanities“ semantisch unterscheiden. J. Taylor hat den Begriff „E-Science“ zutreffend und abstrakt genug wie folgt beschrieben: “eScience is about global collaboration in key areas of science and the next generation of infrastructures that will enable it” (Taylor 2001). In einem Punkt ist diese Beschreibung sehr konkret: Wir werden die Herausforderungen, die er mit dem Begriff „E-Science“ umschreibt, nur bewältigen können, wenn wir eine neue Generation von Infrastrukturen bilden werden. Hiermit sind natürlich nicht nur traditionelle Forschungsinfrastrukturen – wie z.B. große physikalische Einrichtungen wie DESY und CERN – gemeint, sondern vor allem auch neuartige Informationsinfrastrukturen, in denen verteilt vorgehaltene Daten und Werkzeuge in einfacher Weise zusammengeführt werden können.

Der Begriff „*Infrastrukturen*“ umfasst dabei das Bündel an Maßnahmen, das erforderlich ist, um die angestrebten Zielsetzungen effizienter und umfassender als bisher zu erreichen. Hier kann man auf das Beispiel des Internets verweisen, das nichts anderes ist als ein Bündel aus einem Basis-Datenmodell (Daten-Paketen), Hardware-Komponenten wie z.B. Routern, deren Zugriffsmethoden und Protokolle² standardisiert wurden, gepflegten Registrierturen, speziellen Leitungen und einer nicht limitierten Anzahl von Applikationen, die von den standardisierten Mechanismen Gebrauch machen. Das Internet normiert and unterstützt den Austausch von Informationen in unseren Netzwerken. Schon immer konnten Informationen ausgetauscht werden, allerdings waren diese historisch gesehen nicht digitaler Art und

1 Zu einer frühen Beschreibung von E-Humanities vgl. Aschenbrenner et al. (2007).

2 Wie TCP (Transmission Control Protocol), IP (Internet Protocol), SMTP (Simple Mail Transfer Protocol) oder http (Hypertext Transfer Protocol) – alle diese Protokolle werden von der IETF (www.ietf.org) standardisiert.

ihre Überbringung erfolgte verhältnismäßig langsam. Das Internet mit all seinen Mechanismen erlaubt es uns, innerhalb von Sekunden Informationen (z.B. E-Mails) zu einer Vielzahl von Adressaten weltweit zu versenden.

Eine *Informationsinfrastruktur* muss über die erforderlichen Mechanismen verfügen, um in einem verteilten Szenario schnell geeignete Daten und Werkzeuge finden und auf sie zugreifen zu können, damit die Daten interpretiert und in beliebigen Kontexten wiederverwendet werden können. Für jeden datenorientiert arbeitenden Wissenschaftler ist es eine alltägliche Erfahrung, dass all diese Schritte gegenwärtig mit einem erheblichen Aufwand und Zeitverlust verbunden sind. Nur allzu oft verhindert dieser Aufwand, dass interessante Forschungsfragen in Angriff genommen werden, und er führte in den vergangenen Jahren zu Ungleichgewichten. Wissenschafts-Einrichtungen bzw. -Abteilungen, die nicht über das entsprechende Personal mit Spezialkenntnissen verfügen, werden von der daten-orientierten Forschung sogar ausgeschlossen. Dies war und ist einer der wesentlichen Gründe, warum zunehmend E-Humanities-Zentren im akademischen Kontext gebildet wurden und werden – das Spezialwissen muss gebündelt werden, sodass es möglichst vielen Akteuren in der GSW zur Verfügung steht.

Anders als in vielen Fällen in anderen Wissenschaftsdomänen ist die GSW durch einen extrem verteilten *Erzeugungs- und Nutzungsprozess* gekennzeichnet, der für Werkzeuge, aber insbesondere auch für Daten gilt. Hier sind zum einen kommerzielle und akademische Digitalisierungsprojekte zu nennen. Als Resultat der technischen Innovation kann heute aber auch jeder einzelne Wissenschaftler in den GSW mit einem Aufnahmegerät in einfacher Art und Weise relevante Primärdaten erzeugen. Dies gilt vor allem für Audio- und Videoaufnahmen – aber auch die Aufnahme von Augenbewegungen, Gestiken, Hirnpotenzialen und anderem mehr ist heute wesentlich kostengünstiger geworden, sodass viel mehr Labore von diesen Technologien Gebrauch machen können. Auch die Aufnahme und Analyse von funktionellen Magnetresonanztomografien (fMRI) und Gensequenzen hat heutzutage in einigen Instituten der GSW Einzug gehalten. Der neue Trend hin zum Crowdsourcing – der direkten Einbindung einer sehr großen Zahl von Internet-Nutzern in den Prozess des Erzeugens wissenschaftlicher Daten, zunehmend mittels mobiler Geräte – wird die verteilte Generierung von Daten noch verstärken. Wenn wir nicht frühzeitig adäquate Infrastrukturen bereitstellen, die nachhaltig so erzeugte Daten für größere Nutzergruppen vorhalten, können sie nur von wenigen zu der jeweiligen Community gehörenden Personen genutzt werden.

Konsequenz dieses distribuierten Erzeugungsprozesses in den GSW ist eine enorme *Fragmentation* in allen Dimensionen, die geeignete Informationsinfrastrukturen in den GSW noch essenzieller machen als in anderen Wissenschaftsbereichen.

Problemanalyse

Wenn wir die Eckpfeiler einer (technischen) Informationsinfrastruktur für die GSW ausarbeiten wollen, bedarf es zunächst einer Analyse der wesentlichen Trends und Probleme.

Ähnlich wie in den anderen Wissenschaftsdisziplinen können wir eine Reihe von *Trends* beobachten:

- Die verwendeten Daten³ sind immer mehr digitaler Natur mit all den Vorteilen und Nachteilen, die die Möglichkeit der Trennung von Medium und Information mit sich bringt.
- Die Datenmengen nehmen extrem zu, was auch in den GSW eine Folge der wachsenden technologischen Möglichkeiten ist.
- Die Komplexität der Daten – sowohl extern als auch intern – nimmt erheblich zu. Die verschiedenen Datenobjekte stehen in vielfältigen Beziehungen zu anderen Datenobjekten und die Art der Relationen hängt sehr stark von der jeweiligen Nutzung ab, d.h. sie sind nicht statisch. Die Zahl der gespeicherten Informations-Typen nimmt ebenfalls zu, d.h. die interne Struktur und die in diesen Strukturen zum Einsatz kommenden Semantik-Domänen haben ein immer größer werdendes Spektrum.
- Der Zugriff auf Daten vollzieht sich immer öfter in einem anonymen Raum. Während es früher einen direkten persönlichen Austausch von Daten zwischen Erzeugern und Nutzern gab, wobei durch die persönliche Bekanntschaft ein Vertrauensverhältnis auch in die Daten gegeben war, werden in der Zukunft immer mehr Daten verwendet, die von im Web verfügbaren Repositorien zur Verfügung gestellt werden und bei denen allenfalls die Metadaten⁴ Auskunft über den Erzeuger geben. Das Ver-

3 Im Folgenden verweisen wir hauptsächlich auf den Umgang mit Daten, obwohl auch der entsprechende Umgang mit Werkzeugen von großer Bedeutung ist. Die erforderlichen Methoden sind allerdings noch nicht so klar erkennbar.

4 Der Begriff „Metadaten“ wird vielfältig verwendet. In diesem Beitrag werden unter Metadaten Eigenschaft-Werte-Paare verstanden, die die Eigenschaften von Daten-Objekten beschreiben – ähnlich wie eine Karteikarte ein Buch beschreibt.

trauensverhältnis, das verschiedene Aspekte umfasst, muss also durch neue Mechanismen realisiert werden.

- Die einzelnen Wissenschaftler und auch Projektteams sind mit dem gezielten Datenmanagement immer mehr überfordert und viele der erzeugten und von den Wissenschaftlern selbst gespeicherten Daten sind schwer auffindbar und stehen nicht langfristig zur Verfügung.
- Die Wissenschafts-Organisationen sind immer weniger geneigt, große Geldsummen für die Datenerzeugung auszugeben, ohne die Sicherheit zu haben, dass diese Daten vernünftig gepflegt und für andere – und vermehrt auch für die Allgemeinheit – zugänglich sind. Von jedem Wissenschaftler bzw. Projekt, das Daten erzeugt (sei es durch Sensoren, mobile Geräte oder Software) wird erwartet, dass ein Datenmanagement-Plan aufgestellt wird, der Maßnahmen zur Identifikations- und Integritäts-Sicherung, zur Kontext- und Prozess-Beschreibung und zur Langzeit-Sicherung enthält.
- Die Erkenntnis greift um sich, dass sich die Wissenschaft einer neuen Kategorie von Experten bedienen muss. Der HLEG-Report „Riding the Wave“ (2010) nennt diese Experten „Data Scientists“. Sie sind eine neue Art von Bibliothekaren, die in allen Phasen des Umgangs mit digitalen Daten und Werkzeugen ihr Expertenwissen einbringen.

Informations-Infrastrukturen müssen diese Trends antizipieren, d.h. auch, dass sie langfristig angelegt sein müssen. Es bedarf einer Strukturierung unserer heutigen „chaotischen“ Daten-Landschaft, die – ähnlich dem Internet – Modelle, Komponenten, Schnittstellen und Protokolle beschreibt, aber keine Technologien festlegt, denn die werden sich rasant verändern.

Bezüglich der Analyse der wesentlichen Aufgaben wurden in Zusammenarbeit mit L. Lannom (2012) zwei typische kanonische Szenarien durchgespielt. Das erste Szenario beschreibt die prinzipiellen Schritte, die erforderlich sind, wenn man auf Daten zugreifen will (siehe Abb. 1). In diesem Szenario nehmen wir an, dass Nutzer zur Bearbeitung ihrer wissenschaftlichen Fragestellung zunächst mittels geeigneter Eigenschaften eine Suche im Metadaten-Raum starten. Diese wird in einer Anzahl von Datenobjekten, Kollektionen, Werkzeugen oder Methoden resultieren, deren Metadaten zunächst dazu verwendet werden, um zu prüfen, ob die angezeigten Entitäten wirklich den Wünschen der Nutzer entsprechen. Teil dieser Überprüfung wird die Frage sein, ob die Erzeugungskette aus Erzeuger, Kurator und Anbieter als vertrauensvoll angesehen werden kann. Für eine Selektion werden die persistenten Identifikatoren aus den Metadaten entnommen und eventuell wird eine

neue virtuelle Kollektion definiert, die die vorgenommene Selektion umfasst und einen Namen (sowie andere Eigenschaften) erhält. Nunmehr wird der Nutzer auf die Daten bzw. Werkzeuge selbst zugreifen wollen. Dazu müssen die persistenten Identifikatoren in Speicheradressen (URLs) umgesetzt werden, es muss geschaut werden, ob die angezeigten Daten auch nicht manipuliert worden sind, ob der Nutzer eine Zugangsberechtigung bezüglich der selektierten Daten hat, und er muss ein geeignetes Werkzeug haben, um die Daten dann lesen zu können. Letzteres kann ein Herunterladen der Daten beinhalten, aber auch das direkte Sichtbarmachen mittels Web-Applikationen oder das Rechnen auf den Daten z.B. mittels geeigneter Web-Services. Um sinnvoll mit den Daten umgehen zu können, muss der Nutzer die Daten interpretieren können, wozu er im Allgemeinen Struktur- und Semantik-Informationen verwenden muss, die über die Metadaten erreichbar sind. Im Falle einer Nutzung für andere Zielsetzungen wird der Nutzer auch weitergehende kontextuelle Informationen verwenden müssen, die ebenfalls zumindest zu einem Teil (und sei es als Zeiger auf externe Informationen) in den Metadaten gespeichert werden. Schließlich will der Nutzer die Daten auch annotieren, wodurch wieder neue (Meta-) Daten entstehen.

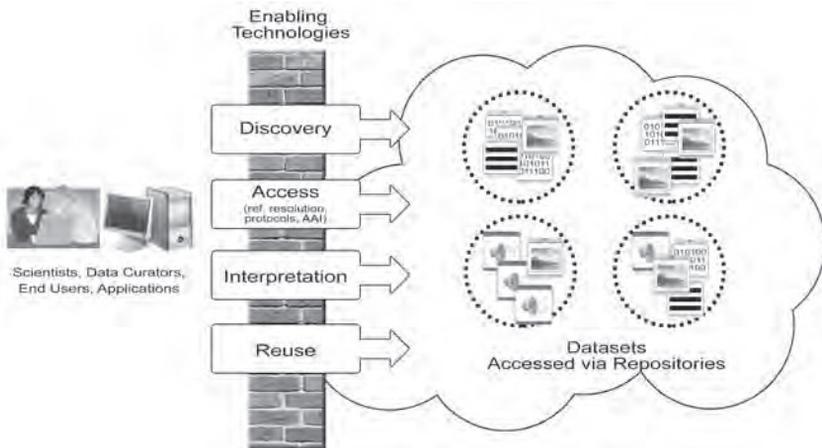


Abb. 1 Grundsätzliche Schritte für Datenzugriff

Ein weiteres typisches Szenario wird in Abbildung 2 beschrieben. Es zeigt einen kanonischen Management-Prozess, in dem der Manager zunächst mittels bestimmter Eigenschaften, die in den Metadaten enthalten sind (z.B. alle Videos mit dem Format MPEG2), eine Selektion vornimmt und diese wahr-

scheinlich zu Dokumentations-Zwecken als eine Kollektion speichert. Darauf erfolgen die Zugriffs-Prozesse, wie sie im ersten Szenario beschrieben wurden. Daran anschließend wird der Manager formalisierte und dokumentierte Kurationsprozesse, z. B. mittels sogenannter „Policy-Regeln“ der Software iRODS⁵ ausführen, um etwa eine Transcodierung einer MPEG2-Video datei in eine H.264/MPEG4-Video datei vorzunehmen, die dann ebenfalls mit einem persistenten Identifier und ergänzenden Metadaten identifiziert und beschrieben wird. Im Anschluss daran wird der Manager schließlich prüfen, ob die Operationen erfolgreich waren.

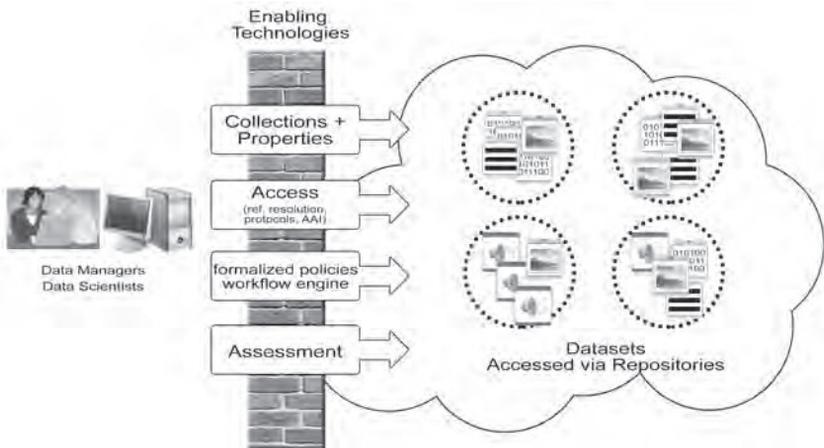


Abb. 2 Daten-Management-Prozess

Weitere derartige typische Szenarien können durchgespielt werden. Sie haben den Vorteil, dass man in einfacher Weise die verschiedenen Schritte identifizieren und damit auch die Mechanismen bestimmen kann, die erforderlich sind, wollen wir in einer strukturierten Weise mit Daten umgehen.

Aus derartigen Szenarien können wir eine Reihe von Schlussfolgerungen ziehen, die sowohl für Datenobjekte als auch Kollektionen gelten, die referenzierbar und damit wiederverwendbar sein sollen:

- Sie brauchen eine über einen Identifikator referenzierbare Identität und mit ihr eng verbunden sind Eigenschaften, die es ermöglichen, z.B. Integrität und Authentizität zu überprüfen,

⁵ <https://www.irods.org/>; http://www.slideshare.net/asist_org/policy-based-data-management-irods-reagan-moore-rdap12

- sie müssen in Repositorien persistent gespeichert und über regelmäßig überprüfte Kurationsprozesse gepflegt werden,
- sie müssen jederzeit über den Identifikator zugreifbar sein,
- sie müssen über Zugriffsregeln geschützt sein, sodass auch in diesem verteilten Szenario nur autorisierte Nutzer Zugriff erhalten,
- sie müssen mittels Metadaten beschrieben sein, die einen Verweis auf den Identifikator enthalten sowie Informationen, die:
 - das Wiederauffinden ermöglichen,
 - die Historie eines Objektes beschreiben,
 - eine Interpretation und Wiederverwendbarkeit erlauben,
- sie müssen mit Zugriffsrechten verbunden sein, die im Allgemeinen für alle Kopien im Internet gleich gehalten werden müssen.

Hieraus können wir die Mechanismen einer technischen Infrastruktur ableiten. Wegen der Langfristigkeit von Forschungsinfrastrukturen muss jeweils geschaut werden, dass alle Mechanismen auf breiten Absprachen, wenn nicht gar internationalen Standards, beruhen.

Technische Infrastruktur

Eine Informationsinfrastruktur muss verschiedene Dimensionen umfassen, wie z.B.

- eine auf Persistenz angelegte organisatorische Struktur, die sie in die Wissenslandschaft einbettet und aus welcher heraus sie gesteuert wird,
- die Ausarbeitung legaler und ethischer Grundsätze, die in konkrete Vereinbarungen und Verträge mündet,
- die Überzeugungsarbeit in den Wissenschaften, um die gegenwärtige Kultur des Umganges mit den Daten entscheidend zu verändern,
- das Ausbilden von „*Data Scientists*“, also einer neuen Generation von Experten, die den einzelnen Wissenschaftlern die Arbeit mit Daten weitgehend erleichtern,
- das Herausbilden einer technischen Infrastruktur, die – ähnlich wie die neuen Trassen, die neuen Brücken und die neuen Signalanlagen den Betrieb von Hochgeschwindigkeitszügen ermöglichen – das reibungslose Arbeiten mit den Daten und auch Werkzeugen ermöglicht.

Im Folgenden werden wir uns auf die Diskussion einiger wichtiger Aspekte der technischen Infrastruktur beschränken.

Landschafts-Strukturierung

Aus dem Vorhergesagten folgt, dass wir eine neue Strukturierung der Datenlandschaft benötigen, in etwa wie es im folgenden Diagramm schematisch angedeutet wird.

Abbildung 3 zeigt links das traditionelle Netzwerk von vielen zufälligen und oftmals temporären und persönlichen Beziehungen, die den Datenfluss regeln. Rechts wird eine Landschaft angedeutet, die durch einige persistente Zentren und stabile Beziehungen bestimmt wird.

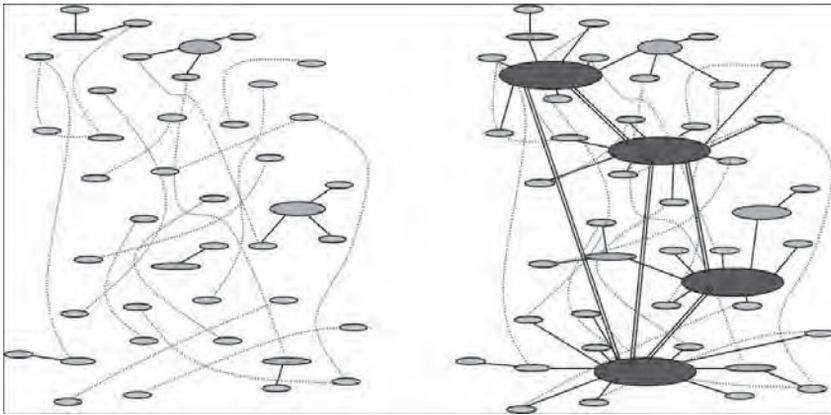


Abb. 3 Netzwerke ohne Zentren und mit Zentren

In einer solchen Landschaft, wie sie im Übrigen momentan von nahezu allen ESFRI-Forschungsinfrastrukturen disziplin-übergreifend etabliert wird, übernehmen einige Zentren die Aufgaben, a) allgemein verfügbare Infrastrukturkomponenten wie z.B. bestimmte Registraturen zu betreiben und anzubieten, b) eine Domäne registrierter Daten und Werkzeuge⁶ zu pflegen, deren Persistenz zu garantieren und den Zugriff zu gewähren, und c) zusammen mit den betreffenden Wissenschaftlern und Software-Entwicklern für

6 Hier sollte hinzugefügt werden, dass die Frage, wie Werkzeuge zentral und persistent angeboten werden können, praktisch keineswegs gelöst ist. „Smarte“ Algorithmen werden weiterhin in den wissenschaftlichen Abteilungen erzeugt und im Allgemeinen müssen die Entwickler auch die Pflege des Codes übernehmen. Es ist noch nicht deutlich, wie trotzdem eine allgemeine und längerfristige Verfügbarkeit abgesichert werden kann. Bezüglich der Daten ist evident, dass die Kuration ein Zusammenspiel mehrerer Faktoren ist, wobei Standards und Best-Practices eine wichtige Rolle spielen.

die Umsetzung von Best-Practices, Standards und kanonischen Arbeitsabläufen zu sorgen. Entscheidend für die Akzeptanz und das Funktionieren einer solchen geänderten Landschaft ist, dass neue Vertrauensverhältnisse etabliert werden:

- Der erzeugende Wissenschaftler muss sich darauf verlassen, dass das Zentrum sorgfältig mit den Daten umgeht.
- Der nutzende Wissenschaftler muss sich darauf verlassen können, dass das Zentrum die Daten mit einer hohen Verfügbarkeit und Persistenz anbietet und dafür Sorge trägt, dass die Daten nicht manipuliert werden.
- Das Zentrum muss sich darauf verlassen, dass sich Erzeuger wie auch Nutzer an gemeinsam erstellte ethische und legale Richtlinien halten.

Basis-Datenmodell

In der durch eine hohe Dynamik geprägten Wissenschaft wird es immer einen Bereich temporärer Daten geben, die aus einer Phase des Testens und Experimentierens entstehen. Die Frage nach dem Moment des Überganges der Daten aus einem „chaotischen“ Datenhinterhof in die Domäne der registrierten und damit öffentlich zugänglichen Daten wird von jedem Institut, von jedem Projekt beurteilt werden müssen. Es scheint angemessen, Software derart zu entwickeln, dass sofort bestimmte Regeln eingehalten werden, sodass es sofort möglich ist, erzeugte Daten automatisch in die Domäne der registrierten Daten zu integrieren – und zwar, ohne dass für den Wissenschaftler ein Mehraufwand entsteht. Für alle Sensoren (Foto-Kameras, fMRI-Scanner, etc.) ist es heute üblich, dass z. B. automatisch Metadaten erzeugt werden.

Um dies zu gewährleisten, benötigen wir ein Basis-Datenmodell, das von allen Beteiligten umgesetzt wird. Wie in Abbildung 4 angedeutet, besteht ein Datenobjekt aus einer Sequenz von Bits, die an verschiedenen Orten gespeichert werden können, einem persistenten Identifikator (PID), der bei einem externen und autorisierten Zentrum registriert wird, und einer Ursprungs-Metadaten-Beschreibung. Metadaten-Beschreibung⁷ und PID verweisen gegenseitig aufeinander, um jederzeit – je nach Zugriffsmethode – die relevan-

⁷ Metadaten-Beschreibungen sind ebenfalls Datenobjekte, d.h. sie haben einen persistenten Identifikator. Da Metadaten offen sind und im Allgemeinen nur wenige Bytes umfassen, muss allerdings davon ausgegangen werden, dass sie in vielfältiger Weise wiederverwendet (kopiert, ergänzt, in Services eingebunden, etc.) werden.

ten Informationen schnell verfügbar zu haben. Mit den PIDs sind Zeiger (z.B. URLs) auf die verschiedenen Speicherplätze der Bit-Sequenzen und andere essenzielle, die Identität der Sequenzen betreffende Attribute gespeichert, wie z.B. eine Prüfsumme. Die Metadaten enthalten, wie bereits erwähnt, weitere Informationen über die Bit-Sequenzen, die wichtig für das Wiederauffinden und die Wiederverwendbarkeit sind. Bezüglich der Attribute kann noch festgehalten werden, dass es disziplinunabhängige Attribute (externe) gibt, die z.B. lediglich das Äußere eines Objektes beschreiben (also technische Metadaten wie Größe in Bytes, Datentyp etc., sowie administrative Metadaten wie Erzeuger, Rechte etc.) und sich somit für disziplinübergreifende Dienste anbieten, und disziplin-spezifische Attribute (interne), also strukturelle Metadaten wie Text- und Dokumentenstruktur sowie deskriptive Metadaten, wie bibliografische oder sonstige inhaltsbezogene Erschließung, die oftmals durch disziplin-spezifische Dienste ausgewertet werden.

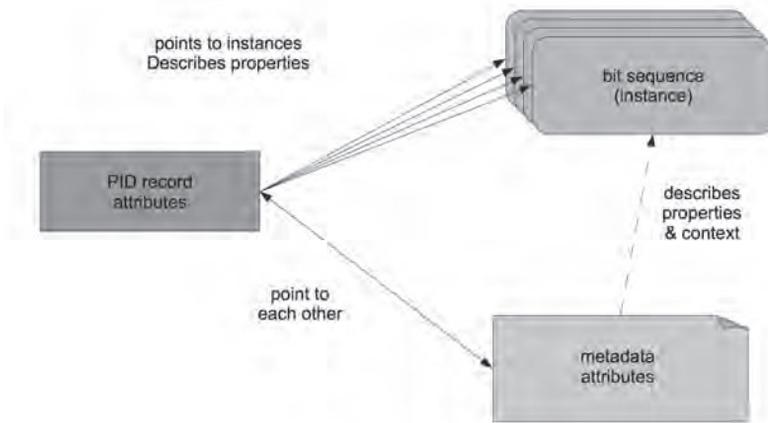


Abb. 4 Relationen zwischen Daten, PID und Metadaten

In jeder Wissenschaftsdisziplin ist der Begriff der Kollektion von großer Bedeutung, insofern er einen engen Zusammenhang zwischen den enthaltenen Datenobjekten darstellt. Kollektionen werden jedoch je nach Verwendungszweck immer neu gebildet – wir sprechen dann gerne über virtuelle Kollektionen, bei denen die einzelnen Objekte sogar in verschiedenen Repositorien gespeichert sein können:

- Die Erzeuger von Daten werden eine Kollektion derart definieren, dass der Erzeugungs-Zusammenhang deutlich ist, der sich zum Beispiel in der Namensgebung ausdrückt: das Korpus des Mittelhochdeutschen, das

Korpus der Musik Mozarts, etc. Dabei kann ein solches Korpus, das zu-
meist nach einheitlichen Regeln erzeugt wird, verschiedene Datentypen
(Texte, Audio-Aufnahmen, Video-Aufnahmen, Zeitreihen, etc.) beinhalten.
Derartige Korpora eignen sich in hervorragender Weise für Zitationen
und sind oftmals der Ausgangspunkt wissenschaftlicher Arbeiten.

- Die Manager von Daten werden vorzugsweise Objekte des gleichen Typs
(alle MPEG2-Videodaten) zu virtuellen Kollektionen zusammenfassen,
um darauf gleichartige Kurations-Prozeduren wie z.B. Transformationen
vornehmen zu können.
- Datennutzer werden wiederum nach eigenen Gesichtspunkten virtuelle
Kollektionen zusammenstellen, die für ihre jeweiligen wissenschaftlichen
Zielsetzungen optimal erscheinen. So könnte z.B. ein Linguist, der
an der Sprach-Entwicklung bei Kindern interessiert ist, aus verschiedenen
Korpora genau die Datenobjekte selektieren, die es erlauben, zwei-
jährige mit vier- und sechsjährigen zu vergleichen. Seine virtuellen Kol-
lektionen wären hierarchisch angeordnet, indem die Altersstufen jeweils
in eigenen Sub-Kollektionen angeordnet wären.

Es mag noch andere Gründe geben, derartige virtuelle Kollektionen zusammen-
zustellen. Im Allgemeinen können wir mithin ein einfaches Datenmodell
für derartige (virtuelle) Kollektionen angeben. Auch eine Kollektion muss
durch Metadaten beschrieben werden, die durch eine PID referenzierbar ge-
macht werden, wie in Abbildung 5 angedeutet ist. Diese Metadatenbeschrei-
bung verweist auf die vielen Metadatenbeschreibungen der in ihr enthaltenen
Sub-Kollektionen und Datenobjekte. Wir müssen auch hier dazu übergehen,
dass Kollektionen extern registriert und damit zitierbar gemacht werden.

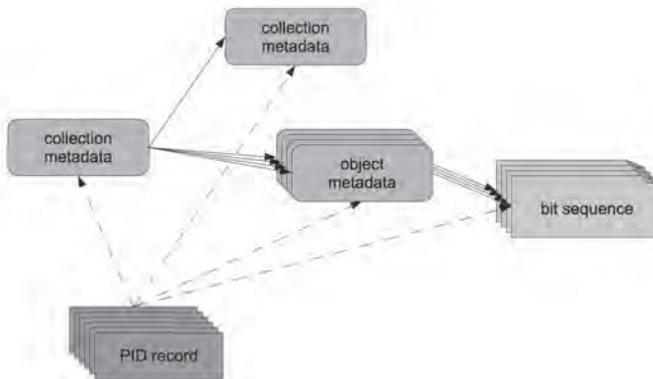


Abb. 5 Relationen zwischen Kollektionen, PID und Metadaten

Natürlich sind wir vor allem in den GSW weit davon entfernt, die Werkzeuge und Infrastrukturen zur Verfügung zu haben, die reibungslos für eine derartige Landschaft von registrierten Datenobjekten und Kollektionen sorgen, sodass alles in einfacher Weise auffindbar, verfügbar, zitierbar und interpretierbar wird. Eine Informationsinfrastruktur muss jedoch dafür sorgen, dass dies mittelfristig möglich wird und auch die bestehenden Daten und Kollektionen in diese Domäne eingebracht werden. Erste Ansätze bestehen allerdings. Für textbasierte Daten wurde z.B. im Rahmen von TextGrid⁸ eine Virtuelle Forschungsumgebung entwickelt, die ein flexibles Metadatenmodell implementiert, das sowohl Datenobjekte (*bit sequence*), und dazugehörige Metadatenobjekte unterstützt als auch Aggregationen (*collection*).⁹ Das gleiche gilt für die flexible CMDI- (Component Metadata Infrastructure) Metadata-Umgebung, die in CLARIN¹⁰ realisiert worden ist.

Registraturen

Für alle verteilten Infrastrukturen sind Registraturen, die von offiziell anerkannten Institutionen gepflegt werden, eine wichtige Voraussetzung. Die Katasterämter, die alle relevanten Informationen über Landparzellen (Eigner, finanzielle Belastungen, chemische Belastungen, darauf befindliche Strommasten, etc.) speichern, mögen ein gutes Beispiel sein. Sie werden vom Staat gepflegt und gehütet und sind für alle entsprechenden Vorhaben die zentrale Anlaufstelle. In ähnlicher Weise regelt das DNS-System¹¹ des Internets mit gepflegten IP-Adressen und Domain-Namen-Registraturen den reibungslosen Austausch der IP-Pakete zwischen registrierten Knoten.

Im Bereich der Daten müssen wir ein ähnliches System der Registraturen aufbauen, die Komponenten wie z.B. die folgenden umfassen und sich über standardisierte Schnittstellen und Protokolle ansprechen lassen:

8 <http://www.textgrid.de>

9 Beschreibungen des TextGrid-Metadatenmodells sowie der Funktionen zur Langzeitarchivierung von Forschungsdaten in den Geisteswissenschaften sind in Pempe (2012) zu finden.

10 <http://www.clarin.eu>

11 Domain-Name-System

- Einrichtungen, die die Registrierung und Resolution von PIDs ermöglichen und die genaue Absprachen bezüglich der zu verarbeitenden Zeichenketten und der mit ihnen gespeicherten Informationen haben,
- Einrichtungen, die Metadaten-Beschreibungen registrieren, aggregieren und für Suchzwecke zur Verfügung stellen,
- Einrichtungen, die Schemata – also Strukturbeschreibungen, die Definitionen semantischer Kategorien, also formalisierter Konzepte – und von vielen verwendete Vokabularien registrieren, aggregieren und zur automatischen Verarbeitung zur Verfügung stellen,
- Einrichtungen, die die Eigenschaften der Zentren und deren Eingangsschnittstellen formal pflegen, um automatische Prozeduren wie Monitoring, Metadatenaggregation etc. ausführen zu können,

Es wird weitere solcher Registraturen geben und zumeist werden diese von einem Netzwerk eng miteinander kollaborierender Einrichtungen betrieben, um so eine hohe Redundanz und somit eine hohe Verfügbarkeit zu bieten. Auch für solche Registraturen gibt es bereits vielfältige Ansätze. So wird z.B. im ESFRI-Projekt DARIAH¹² sowohl eine Collection Registry als auch eine Schema-Registry aufgebaut, die langfristig eine generische Suche über heterogene Kollektionen ermöglichen werden. Ähnliche Aktivitäten gibt es im Schwester-Projekt CLARIN.

Persistente Identifikatoren

Es scheint nunmehr eine breite Überzeugung zu geben, dass es eines der Kernaufgaben für ein sauberes Datenmanagement ist, die Objekte mittels extern registrierter PIDs eindeutig identifizierbar und somit referenzierbar zu machen. Ebenfalls wird weithin akzeptiert, zusammen mit den PIDs verschiedene Informationen zu speichern wie z.B. die Adressen der Speicherorte der Kopien der Bit-Sequenzen und der Ursprungs-Metadaten-Beschreibung, eine Prüfsumme zur Integritätsprüfung, die Ursprungs-Repository, um Eigentümer und Zugriffsrechte auffinden zu können, und dergleichen. Momentan werden im Rahmen der Research Data Alliance¹³ diesbezügliche Harmonisierungen und Standardisierungen vorgenommen.

¹² <http://www.dariah.eu/>

¹³ <http://rd-alliance.org>

Der PID-Service ist ein typisches Beispiel dafür, dass es allgemeine Services gibt, die vollkommen disziplinunabhängig sind. Momentan kann man auf zwei weithin verfügbare entsprechende Dienste verweisen, die beide auf der gleichen Technologie beruhen, jedoch verschiedene Zielsetzungen haben: (1) DataCyte¹⁴ will alle zitierbaren Kollektionen mittels DOIs (Digital Object Identifier¹⁵) registrieren und zugleich mit entsprechenden Beschreibungen anbieten; (2) EPIC (European Persistent Identifier Consortium)¹⁶ richtet sich auf die frühe Registrierung der Millionen Datenobjekte, die in der Wissenschaft täglich erzeugt werden. Beide Angebote basieren auf dem robusten und skalierbaren Handle-System¹⁷. Da die Syntax der Handles auch die Adressierung von Fragmenten eines Objektes (z.B. Teile eines Videos) erlaubt, sehen Handles im Allgemeinen wie folgt aus:

<Handle-Prefix>/<Handle-Suffix>#<Fragment-Identifizier>

Der Handle-Prefix wird global vergeben (zeigt auf die registrierenden Einrichtungen) und aufgelöst, der Handle-Suffix wird durch die registrierende Einrichtungen (z.B. EPIC) vergeben und aufgelöst. Die Fragment-Identifizier können nur durch das Repositorium aufgelöst werden, da z.B. Zeit- und Strukturangaben, die auf das Fragment verweisen, noch nicht standardisiert sind. Es ist die registrierende Einrichtung, die Richtlinien darüber aufstellt, wie ein Suffix aufgebaut werden soll. Im Allgemeinen setzt sich die Erkenntnis durch, dass möglichst wenig Semantik in der Zeichenkette enthalten sein sollte. Allerdings kann, wie im Fall EPIC, ein Handle-Suffix ein Verweis auf die registrierende Einrichtung enthalten, da das Handle-Prefix das gesamte EPIC-Konsortium adressiert.

Metadaten (MD)

MD-Beschreibungen erfüllen verschiedene Zwecke und verschiedene Communities haben eigene Kategoriensätze hinzugefügt und diese mit Namen versehen wie z.B. MD für die Präservierung, für das Auffinden, zur Siche-

14 <http://datacite.org/>

15 Auch DOIs sind Handles, allerdings sind DOIs mit dem durch die IDF/DataCyte definierten Business-Modell verknüpft. EPIC hat ein anderes Business-Modell.

16 <http://www.pidconsortium.eu/>

17 <http://www.handle.net/>

rung der Historie (*Provenance*), für das Management, etc. Letztlich liegt all diesen Initiativen zugrunde, dass die Metadaten-Beschreibungen kontextuelles Wissen in Form von Eigenschaft-Werte-Relationen vorhalten, die bei einer Wieder- bzw. Weiterverwendung des betreffenden Objektes oder der Kollektion für verschiedene Aufgaben helfen. Wir können also vermuten, dass Metadaten-Beschreibungen im Laufe des *Lifecycle* der Daten kontinuierlich ergänzt bzw. angepasst werden. Wir brauchen also einen Mechanismus, der sowohl modular als auch flexibel gegenüber Erweiterungen ist.

Für die Weiterverwendung von Metadaten ist es erforderlich, dass 1.) die ihnen zugrunde liegenden Schemata registriert und verfügbar sind, 2.) die Definitionen der verwendeten Datenkategorien ebenfalls registriert und verfügbar sind und 3.) die Metadaten mittels eines standardisierten Protokolls wie z.B. OAI/PMH angeboten werden. Durch diese drei Maßnahmen wird es Service-Anbietern ermöglicht, Metadaten gezielt zu aggregieren, sinnvolle semantische Brücken zu bauen und spezielle Portale anzubieten. Es gibt eine ganze Reihe von Einrichtungen, die derartige Portale anbieten. Hier sei nur auf zwei Beispiele verwiesen: (1) Europeana¹⁸ möchte ein Portal sein, über das man mittels eines einfachen Zugangs auf europäische Museums- und Archiv-Kollektionen zugreifen kann; (2) das Virtual Language Observatory¹⁹ aggregiert Metadaten über Sprachdaten weltweit. Die Methoden in diesen Portalen sind weitgehend identisch:

- Die MD werden zunächst aggregiert und in einer Datenbank zusammengefasst.
- Mittels Selektionen und Abbildungen wird ein gemeinsamer Index aufgebaut.
- Dieser eignet sich dazu, mittels Abfragen schnell Resultate zu liefern.
- Es wird auch ein Hinweis auf die ursprünglichen Metadaten mitgeliefert, um tiefere Inspektionen ausführen zu können.

Momentan ist bei vielen dieser Portale zu erkennen, dass die von den Anbietern gelieferten Metadaten in vielerlei Hinsicht keinen hohen Qualitäts-Standards entsprechen, was einen hohen Aufwand an teurer manueller Kuration erfordert.

18 <http://www.europeana.eu/portal/>

19 <http://www.clarin.eu/vlo/>

Verteilte Authentifizierung und Autorisierung

Momentan leben wir in einer Zeit, in der noch viele Web-Portale, die geschützte Dienste anbieten, eine Registrierung der Nutzer fordern. Dies führt zu dem bekannten Phänomen, dass jeder Nutzer eine Vielzahl von Namen/Password-Kombinationen pflegen muss, was wiederum nach sich zieht, dass Wege der Memorisierung gefunden werden, die die Kombinationen keineswegs so sicher machen, wie es erforderlich wäre.

Sowohl im akademischen Bereich als auch im Bereich der sozialen Netzwerke etablieren sich aber Technologien, die hier Abhilfe schaffen können. Die in den sozialen Netzwerken verwendeten Technologien (OpenID) sind jedoch für ernsthafte Anwendungen als zu unsicher anzusehen. Zum einen gibt es im Protokoll selbst Sicherheitsprobleme, zum anderen setzt der Erhalt einer OpenID nicht zwingend eine Überprüfung der Identität voraus, sodass eine Rückverfolgung des Benutzers nicht gewährleistet ist.

Da technische Infrastrukturen für die GSW auf Ressourcen beruhen, die von Rechenzentren betrieben werden, deren Nutzungsbedingungen anonyme oder pseudonyme Nutzung ausschließen, und da viele über diese Infrastrukturen bereitgestellte Inhalte Lizenzen unterliegen, die ebenfalls eine Identitätsprüfung voraussetzen, sind sichere Authentifizierungstechnologien eine Voraussetzung. Im akademischen Umfeld hat sich hier die Föderationstechnologie durchgesetzt, die auf dem Standard SAML (Security Assertion Markup Language)²⁰ basiert, nicht zuletzt wegen des Nebeneffekts Single Sign-on, wodurch der Benutzer sich nur einmal am Tag authentifizieren muss und danach für alle in der Föderation zur Verfügung gestellten Dienste authentifiziert ist. Das eine Passwort ist in der sogenannten Heimatorganisation des Benutzers gespeichert, zumeist die eigene Hochschule, also das Passwort, mit dem man sich z.B. im Hochschul-E-Mail-Server authentifiziert.

Alle Attribute, die die Heimatorganisation über den Nutzer vorhält und die die Heimatorganisation gewillt ist, innerhalb der Föderation weiterzugeben, können von den Diensteanbietern für Autorisierungsentscheidungen genutzt werden.

Eine Föderation ist letztendlich ein Vertrauensbund zwischen Benutzerverwaltungen (die *Identity-Provider*) und Dienste-Anbietern (die *Service-Provider*). Über Verträge versichern im Wesentlichen:

20 <http://saml.xml.org/saml-specifications>

- die Identity-Provider, dass die Benutzerverwaltung aktuell ist, also z.B. ein exmatrikulierter Student sich nicht mehr im Rahmen der Föderation authentifizieren kann
- die Service-Provider, dass sie die ihnen von Identity-Providern gelieferten personenbezogenen Daten nur zum Zwecke der Dienstbringung nutzen, also keinen Missbrauch der Daten begehen.

Auf nationaler Ebene haben sich in verschiedenen europäischen Ländern entsprechende Föderationen etabliert, etwa die DFN-AAI²¹ in Deutschland oder die SWITCH-AAI²² in der Schweiz. Innerhalb dieser Produktivdienste können sehr viele Hochschulangehörige über ihren Heimat-Account eine Vielzahl von Diensten nutzen, insbesondere Angebote wissenschaftlicher Verlage, aber auch viele weitere wissenschaftliche und kommerzielle Dienste²³.

Die Grundlagen für eine verteilte Authentifizierungs- und Autorisierungsinfrastruktur (AAI) sind also gelegt und Infrastrukturprojekte in den GSW wie etwa CLARIN und DARIAH bauen auf diese Grundlagen auf. Es gibt jedoch eine Reihe von Herausforderungen, die bewältigt werden müssen, damit diese Infrastrukturen erfolgreich eingesetzt werden können:

- Es gibt noch keine vollständige, über Ländergrenzen hinweg operierende Föderation. Eine europaweite Föderation ist zwar gegenwärtig im Aufbau (eduGAIN²⁴), diese deckt aber gegenwärtig nur einen geringen Teil der für GSW wichtigen Hochschulen ab. Deshalb wird im Rahmen von CLARIN gegenwärtig eine Ad-hoc-Föderation aufgebaut, die auch von DARIAH so lange nachgenutzt wird, bis sich eduGAIN als Lösung anbietet.
- Viele Hochschulen weigern sich, den Service-Providern der im Rahmen der GSW-Infrastrukturen angebotenen Dienste auch nur einen Minimalersatz von Attributen weiterzugeben, meistens mit dem Argument, dass sonst der Datenschutz nicht gewährleistet ist. Hier gibt es verschiedene Ansätze, um zu einer Lösung zu kommen:

21 <https://www.aai.dfn.de/>

22 <http://www.switch.ch/aai/>

23 Vgl. z. B. die wichtigsten in der DFN-AAI angebotenen Dienste unter <https://www.aai.dfn.de/verzeichnis/sp-dfn-aai/>.

24 <http://www.geant.net/service/edugain/>

- Im Rahmen der TERENA-Initiative „Data protection Code of Conduct“²⁵ [21] werden Regeln definiert, die ein Service-Provider befolgen muss, um EU-Datenschutz-konform personenbezogene Daten erhalten zu können. Alle Service-Provider, die sich vertraglich verpflichten, diese Regeln einzuhalten, werden entsprechend markiert, sodass Identity-Provider dieser Kategorie von Service-Providern personenbezogene Attribute liefern können.
- Wenn auch nur ein über Service-Provider hinweg eindeutige Identifikator von den Identity Providern geliefert wird, können die jeweiligen Infrastrukturen weitere Attribute und Berechtigungsgruppen selbst pflegen. Über einen im SAML-Protokoll vorgesehenen Mechanismus können die von der Heimatorganisation und von der sogenannten virtuellen Organisation (z.B. DARIAH) gesammelten Informationen zusammen an den Service-Provider weitergegeben werden. Dieser Mechanismus wurde im Rahmen von DARIAH-DE²⁶ implementiert und erfolgreich getestet.
- Schließlich wird es entsprechende Kommunikation benötigen, um über diese Ansätze zu informieren und datenschutzrechtliche Bedenken auszuräumen. Eine erste gemeinsame CLARIN/DARIAH-Aktivität in dieser Richtung war der „CLARIN-D and DARIAH-D call for action on AAI“²⁷.
- Gegenwärtig wird vorwiegend das Web-SSO-Profil von SAML implementiert, das für Dienste, die als Web-Dienste angeboten werden, funktioniert. Allerdings werden im Rahmen der GSW-Infrastrukturen auch Dienste angeboten, die nicht über den Webbrowser bedient werden. Solche Dienste, wie etwa Web Services, ebenfalls in die AAI einzubinden, erfordert zusätzliche Technologien. Zum einen kann hier das ebenfalls in SAML standardisierte Enhanced Client Proxy Profile verwendet werden, was bereits erfolgreich getestet wurde. Andererseits wurde kürzlich OAuth2 (vgl. Hardt 2012) standardisiert, das hier eingesetzt werden könnte. Gegenwärtig wird im Rahmen von DARIAH evaluiert, wie diese Technologie in die SAML-basierte AAI integriert werden kann.

25 Vgl. https://refeds.terena.org/index.php/Data_protection_coc.

26 <http://de.dariah.eu/>

27 http://www.clarin.eu/sites/default/files/clarin_dariah_call-for-action-aa1.pdf

Schlussbemerkung

Es wurde hier – wie in vielen anderen Dokumenten (vgl. HLEG 2010; van der Graaf/Waaijers 2012) – argumentiert, dass es einer neuartigen Strukturierung des Raumes unserer wissenschaftlichen Daten und auch neuer Werkzeuge bedarf, um es den datenorientiert arbeitenden Wissenschaftlern insbesondere auch in den GSW zu ermöglichen, die vielen Forschungsfragen (große wie auch die vielen „kleinen“) wesentlich effizienter und tiefgründiger angehen zu können, als es heute der Fall ist. Dem Aufbau von Informationsinfrastrukturen kommt dabei eine Kernbedeutung zu, der sich die Wissensorganisationen nicht entziehen können. Die technische Infrastruktur stellt dabei nur einen der verschiedenen Aspekte derartiger Informationsinfrastrukturen dar, die es zu lösen gilt, wobei festgestellt werden kann, dass die erforderlichen organisatorischen und kulturellen Veränderungen, die es zu bewältigen gilt, um ein Vielfaches mehr an Zeit und Aufwand aller Beteiligten erfordern werden. Die technische Infrastruktur wird sich immer wieder einerseits den veränderten Anforderungen und Umgebungen stellen und andererseits der technologischen Innovation Rechnung tragen müssen. Sie muss daher inhärent flexibel angelegt sein und vor allem Standards/Best Practices für Komponenten wie z. B. Registraturen, Formate, Protokolle und Schnittstellen festlegen, sodass immer wieder neue Technologien eingesetzt werden können.

Ebenso muss die technische Infrastruktur eine Strukturierung des Raumes der Daten und Werkzeuge in Richtung auf Zentren fördern, die ähnlich wie die Bibliotheken in früheren Phasen Verantwortung übernehmen. Diese Zentren müssen reine Service-Einrichtungen sein, wobei im Moment nicht abschätzbar ist, welche Art Zentren und wie viele wir benötigen werden und wie die verschiedenen Zentrentypen in dem entstehenden System disziplinübergreifender und disziplinspezifischer Dienste miteinander kollaborieren werden. Sowohl Bibliotheken, Datenzentren sowie neuartige Konstruktionen werden eine Rolle übernehmen, wenn sie sich entsprechend der verändernden Anforderungen weiterentwickeln.

Aufgrund verschiedener Übersichten und Umfragen²⁸ ist zumindest deutlich, dass sich gegenwärtig die Datenorganisationen und mithin die Anforde-

28 Vgl. z. B. die Projekte Radieschen (<http://www.forschungsdaten.org/>) und EUDAT (<http://www.eudat.org>).

rungen an eine technische Infrastruktur in nahezu allen Wissenschaftsdisziplinen grundlegend ändern.

Literaturverzeichnis

- Aschenbrenner, A. et al. (2007): Von e-Science zu e-Humanities – Digital vernetzte Wissenschaft als neuer Arbeits- und Kreativbereich für Kunst und Kultur. In: *Bibliothek* 31 (2007): 1, 11–21. Online: http://www.b2i.de/fileadmin/dokumente/BFP_Bestand_2007/Jg_31-Nr_1/Jg_31-Nr_1_Aufsaeetze/Jg_31-2007-Nr_1-S_11-21.pdf.
- ESFRI (2006): Roadmap 2006. http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri-roadmap§ion=roadmap-2006.
- Hardt, D. (Hrsg.) (2012): The OAuth 2.0 Authorization Framework, RFC 6749, October 2012. <http://tools.ietf.org/html/rfc6749>.
- HLEG (2010): Riding the wave. How Europe can gain from the rising tide of scientific data. Final report of the High level Expert Group on Scientific Data. A submission to the European Commission. October 2010. <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>.
- Lannom, L. (2012): DAITF: Enabling Technologies (Präsentation beim Pre-ICRI DAITF Workshop, Copenhagen, März 2012)
- Liegmann, H. (2007): Einführung. In: Neuroth, H. et al. (Hrsg.): *nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung. Version 0.1*. Göttingen: nestor – Kompetenznetzwerk Langzeitarchivierung digitaler Ressourcen für Deutschland. http://nestor.sub.uni-goettingen.de/handbuch/artikel/nestor_handbuch_artikel_1.pdf.
- Pempe, W. (2012): Geisteswissenschaften. In: Neuroth, H. et al. (Hrsg.): *Langzeitarchivierung von Forschungsdaten – Eine Bestandsaufnahme*. Boizenburg: Hülsbusch, S. 137–159. Online: http://nestor.sub.uni-goettingen.de/bestandsaufnahme/kapitel/nestor_bestandsaufnahme_007.pdf.
- Taylor, J. (2001): Präsentation beim UK e-Science Town Meeting (Juli 2001).
- van der Graaf, M.; Maaijers, L. (2012): A Surfboard for Riding the Wave. Towards a Four Country Action Programme on Research Data. A Knowledge Exchange Report. 2. Aufl. [¹2011]. http://www.knowledge-exchange.info/Admin/Public/DWSDownload.aspx?File=%2fFiles%2fFiler%2fdownloads%2fPrimary+Research+Data%2fSurfboard+for+Riding+the+Wave%2fKE_Surfboard_Riding_the_Wave_Screen.pdf.

Ausblick

Die wissenschaftliche Bibliothek im Kontext von Forschungsinfrastrukturen

von Heike Neuroth

Einleitung

Das Verständnis und Aufgabenprofil von wissenschaftlichen Bibliotheken ist seit Jahrhunderten, die Forschung und die Lehre zu unterstützen. Während in der Vergangenheit dazu hauptsächlich die Unterstützung bei der Suche nach analoger Information, die Aufbereitung oder Aufbewahrung von analoger Information dazu gehörte, ist es im digitalen Informations- und Wissenszeitalter damit alleine nicht mehr getan. Durch den rasanten Technologiewechsel,¹ die Unmenge an produzierten digitalen Daten,² der stark zunehmenden nationalen und internationalen Vernetzung der Wissenschaftler zu zum Teil sehr großen Forschungsverbänden³ und der anwachsenden Anzahl (interdisziplinärer), auf digitalen Technologien basierender Kooperations- und Kollaborationskulturen,⁴ die auch mit neuen Kommunikationswegen einhergehen, wird die Rolle und Bedeutung von wissenschaftlichen Bibliotheken vielfältiger und komplexer.

Bleibt die Kernaufgabe von wissenschaftlichen Bibliotheken, nämlich Forschung (und Lehre) und die Forschungsprozesse zu unterstützen, bestehen, so lohnt ein Blick auf einige Wissenschaftsdisziplinen, um zu verstehen, wie sich die Wissenschaftsprozesse in den letzten Jahrzehnten (z.B. Natur- und Lebenswissenschaften) oder letzten Jahren (z.B. Geistes- und Sozialwissenschaften) verändert haben. Während in früheren Jahrhunderten und Jahrzehnten die Wissenschaftler weitestgehend individuell und isoliert geforscht haben, entstehen heutzutage quer über die Fachgebiete hinweg kleinere

1 Vgl. z.B. Hightech Strategie des BMBF, <http://www.hightech-strategie.de/>.

2 Vgl. z.B. EU-Initiative: A Vision for Global Research Data Infrastructures, <http://www.grdi2020.eu/>.

3 Vgl. z.B. BBMRI (Bio banking and Biomolecular Resources Research Infrastructure), <http://www.bbMRI.eu/>.

4 Vgl. z.B. International Astronomical Union (IAU), <http://www.iau.org/about/>.

Teams und größere Forschungsverbände, die nicht vor Landesgrenzen oder Sprachbarrieren haltmachen. In der heutigen Zeit ist es üblich, dass mehrere Wissenschaftler, in einigen Disziplinen wie zum Beispiel in der Teilchenphysik sogar tausende Forscher, auf (gemeinsam erstellte) Messergebnisse zugreifen, sie weiter verarbeiten und zunehmend gemeinsam publizieren. „Data sharing“ und „data re-use“ sind im internationalen Sprachgebrauch kein Novum mehr, auch in den eher traditionell arbeitenden Disziplinen (z.B. in geisteswissenschaftlichen Fächern) setzt sich mehr und mehr die Erkenntnis durch, dass mithilfe digitaler Methoden „alte“ Forschungsergebnisse „neu“ interpretiert werden können und dass mithilfe neuer digitaler Verfahren neue Erkenntnisse gewonnen werden können (Stichwort Digital Humanities⁵ bzw. E-Humanities).

Im Folgenden werden einige aktuelle Beispiele aus unterschiedlichen Fachdisziplinen aufgeführt, um zu verdeutlichen, wie Forschungsprozesse heutzutage mithilfe digitaler Verfahren und Methoden ablaufen (können):⁶

- Die Lebenswissenschaften mit dem DNA-Sequenzierer, einem Verfahren zum automatisierten Ablesen der Sequenzinformation von einem DNA-Molekül: Die Sequenzierungsverfahren der zweiten Generation erlauben heutzutage zum Beispiel die Analyse des menschlichen Genoms in ca. acht Tagen.
- Die Hochenergiephysik mit dem Großgerät Teilchenbeschleuniger in Genf, der pro Jahr mehrere Petabyte an Daten produziert. Es fallen sehr hohe Datenmengen an, die von verschiedenen Forschergruppen verarbeitet werden und daher archiviert werden müssen.
- Die Geowissenschaften mit der Erkundung des Ozeanbodens durch wissenschaftliche Bohrungen, ein großes, international angelegtes, langjähriges Projekt, für dessen beträchtliche Datenmengen eine spezifische Forschungsdateninfrastruktur entwickelt werden musste.
- Die Klimawissenschaften sind zwar eine recht junge Disziplin, aber dafür bereits sehr früh auf internationale Kooperation angelegt, wie zum Beispiel die Berichte des Intergovernmental Panels on Climate Change (IPCC) zeigen.

5 Vgl. z.B. http://de.wikipedia.org/wiki/Digital_Humanities.

6 Diese Liste beruht auf Neuroth (2012).

- Die Astronomie als eine der ältesten Fachwissenschaften produziert nicht nur riesige Datenmengen, sondern benötigt große Rechenleistung zum Beispiel für die Modellierung von Supernova-Explosionen.
- Die Textwissenschaften generieren zwar nicht vergleichbar große Datenmengen und benötigen daher keine ähnlich hohe Rechenleistung, aber ihre Forschungsfragen basieren zunehmend auf umfangreichen digitalen Textsammlungen.⁷ Data Mining oder Text Mining sind hier die Schlagwörter der Gegenwart und Zukunft.
- Die Psycholinguistik mit dem Weltatlas der bedrohten Sprachen, für den große Datenmengen durch die Video- und Tonaufzeichnungen anfallen, stellt ein weiteres Beispiel einer relativ jungen Fachdisziplin dar.

Gemeinsam ist allen hier aufgeführten Beispielen, dass die Forschung in einer vernetzten Umgebung im Rahmen größerer Kooperationsverbünde stattfindet und die Forschungsprozesse IT-basiert ablaufen. Dies gilt sicherlich nicht für alle Wissenschaftsdisziplinen gleichermaßen und auch nicht für alle Forschungsprozesse. Allerdings dürfte sich dieser Trend eher noch fortsetzen und ausweiten, sodass auch in Zukunft der Einsatz von Technologien, digitalen Verfahren und Methoden sowie die Produktion, Veröffentlichung und Nachnutzung von digitalen Daten deutlich zunehmen wird. Einher damit geht auch, dass die Forschungsprozesse mehr und mehr in einen virtuellen Raum verlagert werden, sodass z.B. mehrere Wissenschaftler aus unterschiedlichen Ländern auf große und teure Messinstrumente (z.B. Teilchenbeschleuniger, Radio-Teleskop, Massenspektrometer etc.) zugreifen können, ohne diese vor Ort bedienen zu müssen. Auch digitale Werkzeuge und Dienste werden verstärkt in internationalen Kooperationsverbänden entwickelt und anderen Forschergruppen für ihre Daten-Prozessierung (kostenfrei) zur Verfügung gestellt. Ein bekanntes kommerzielles Beispiel stellt der von Google entwickelte Ngram Viewer⁸ dar, ein Analysetool für Data-Mining-Verfahren in größeren Textmengen. Ein länderübergreifender Ansatz, um den vertrauensvollen Zugang zu Virtuellen Forschungsumgebungen und Forschungsinfrastrukturen zu garantieren, stellt die sog. Authentifizierungs- und Autorisierungs-Infrastruktur AAI⁹ dar, welche bei einmaliger Zugangsvergabe

7 Vgl. z.B. vgl. das Forschungsprojekt Codex Sinaiticus, <http://codexsinaiticus.org/de/>.

8 <http://books.google.com/ngrams>

9 Vgl. z.B. <https://www.aai.dfn.de/>.

(account) den dezidierten Zugang zu (lizensierten) Textcorpora oder Daten, Werkzeugen und Diensten erlaubt.

Welche Rollen können und werden wissenschaftliche Bibliotheken in diesem Kontext einnehmen, welche neuen Aufgaben werden sie übernehmen? Und wie müssen sie sich neu aufstellen und ausrichten, um bei diesen Veränderungsprozessen nicht ihre Daseinsberechtigung zu verlieren oder auf ihre Rolle als „Bewahrer gedruckter Werke“ reduziert zu werden?

Aktuelle Initiativen, Strategiepapiere, Diskussionen

Seit geraumer Zeit werden zahlreiche Strategiepapiere und Empfehlungen von Wissenschaftsorganisationen, Förderern und politischen Gremien veröffentlicht, die eine Neuorganisation und Neupositionierung der relevanten Akteure im Bereich der Informations- und Wissenschaftsgesellschaft fordern. Im Folgenden werden einige für wissenschaftliche Bibliotheken relevante Publikationen genannt:

- Die 2010 von der *Deutschen Forschungsgemeinschaft* (DFG) veröffentlichten Empfehlungen der Kommission für IT-Infrastruktur für 2011–2015 „Informationsverarbeitung an Hochschulen – Organisation, Dienste und Systeme“ (DFG 2010) zielen zwar hauptsächlich auf das Aufgabenspektrum von wissenschaftlichen Rechenzentren, in der Zusammenarbeit mit wissenschaftlichen Bibliotheken ergeben sich hieraus allerdings auch interessante Impulse und neue Wege. Konkret auf die zukünftige Rolle und Aufgaben von wissenschaftlichen Bibliotheken geht das im Jahr 2012 erschienene Positionspapier der DFG, erarbeitet vom Ausschuss für wissenschaftliche Bibliotheken und Informationssysteme, mit dem Titel „Die digitale Transformation weiter gestalten“ (DFG 2012a) ein.
- Die *Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen* (Allianzinitiative 2013) verabschiedet kontinuierlich konkrete, zum Teil operationale Lösungen und Empfehlungen zu bestimmten Handlungsfeldern wie „Open Access“, „Nationale Lizenzierung“, „Forschungsdaten“ oder „Virtuelle Forschungsumgebun-

gen“. Die Allianz-Initiative startet mit Beginn 2013 in ihre zweite fünfjährige Phase.¹⁰

- Die von der *Gemeinsamen Wissenschaftskonferenz des Bundes und der Länder* (GWK) angestoßene Kommission „Zukunft der Informationsinfrastruktur“ (KII)¹¹ veröffentlichte im April 2011 ein „Gesamtkonzept für die Informationsinfrastruktur in Deutschland“.
- Der *Wissenschaftsrat* (WR) hat im Jahr 2011 eine Reihe weitreichender Empfehlungen im Bereich Informationsinfrastrukturen veröffentlicht (vgl. Wissenschaftsrat 2011). Im Jahr 2012 konkretisieren die publizierten „Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020“ (Wissenschaftsrat 2012) die Handlungsfelder, insbesondere im Hinblick auf die Rolle der wissenschaftlichen Bibliotheken. Lesenswert zu dieser WR-Empfehlung dürfte auch die DFG-Stellungnahme (vgl. DFG 2012b) sein.
- Die *Hochschulrektorenkonferenz* (HRK) hat Ende 2012 die Empfehlung „Hochschule im digitalen Zeitalter: Informationskompetenz neu begreifen – Prozesse anders steuern“ (HRK 2012) veröffentlicht. Auch in dieser Empfehlung wird den wissenschaftlichen Bibliotheken eine zentrale Bedeutung angetragen und auf neue Aufgabenfelder aufmerksam gemacht.
- Die *Enquete-Kommission des Deutschen Bundestages* „Internet und digitale Gesellschaft“¹² hat im Rahmen der „Projektgruppe Bildung und Forschung“ eine Reihe von Handlungsempfehlungen (vgl. Deutscher Bundestag 2012) veröffentlicht, die einen deutlichen Bezug zu dem Aufgaben- und Kompetenzspektrum wissenschaftlicher Bibliotheken darstellen. Dies verwundert nicht, da zu den sogenannten Expertenhearings (vgl. Deutscher Bundestag 2011) sowohl Vertreter von wissenschaftlichen Bibliotheken als auch Förderern (DFG) eingeladen waren.
- Auf *EU-Ebene* ist es nahezu unmöglich, auch nur eine treffende Auswahl relevanter Veröffentlichungen zu zitieren. Der 2010 veröffentlichte Bericht „Riding the wave – How Europe can gain from the rising tide of scientific data. Final report of the High Level Expert Group on Scientific

10 Vgl. Pressemitteilung vom 17. Januar 2013, http://www.allianzinitiative.de/de/steuerungsgremium/allianz_der_wissenschaftsorganisationen_setzt_die_schwerpunktinitiative_digitale_information_fort/.

11 <http://www.leibniz-gemeinschaft.de/?nid=infrastr>

12 <http://www.bundestag.de/internetenquete/>

Data. A submission to the European Commission“ (EU 2010) weckte sicherlich zum ersten Mal das Bewusstsein, dass im Forschungskontext in naher Zukunft mit ganz anderen Herausforderungen zu rechnen sein wird. Auch wenn dies vielleicht in absehbarer Zeit nicht alle Wissenschaftsdisziplinen gleichermaßen betreffen wird, so wird doch deutlich, vor welchen (neuen) Aufgaben Service-Infrastruktureinrichtungen wie wissenschaftliche Rechenzentren und Bibliotheken stehen werden. Der Begriff Forschungsinfrastrukturen ist auf europäischer und dann auch auf deutscher Ebene¹³ durch das EU-ESFRI-Programm („European Strategy Forum on Research Infrastructures“¹⁴) eingeführt worden. Der Begriff Forschungsinfrastrukturen ist nicht neu und insbesondere in den Naturwissenschaften seit Jahrzehnten für die Konzeption, die Entwicklung und den Betrieb analoger Forschungsinfrastrukturen etabliert (z.B. CERN¹⁵ für die Teilchenphysik, Polarstern¹⁶ für die Umwelt- und Klimaforschung etc.). Neu seit ca. zehn Jahren ist die Übertragung des Konzeptes auf die Konzeption, die Entwicklung und den Betrieb für digitale Forschungsinfrastrukturen und die explizite Berücksichtigung geistes- und sozialwissenschaftlicher Disziplinen. Es ist jetzt schon deutlich geworden, dass im neuen EU-Rahmenprogramm „Horizon 2020“¹⁷ mit Start Ende 2013 digitale Forschungsinfrastrukturen eine bedeutsame Rolle spielen werden.

Die Fülle an Positionspapieren, Empfehlungsberichten etc. quer über die politischen Gremien und förderpolitisch relevanten Einrichtungen hinweg macht deutlich, dass Deutschland vor einer Neuausrichtung der Informationsinfrastruktur bzw. Forschungsinfrastruktur steht, die insbesondere digitale Transformationsprozesse in der Forschung und Lehre berücksichtigt. Auch ist klar geworden, dass bestimmte Erkenntnisse die politische Ebene – seien es der Bund oder einzelne Bundesländer¹⁸ – bereits erreicht haben. Inwieweit

13 Im Koalitionsvertrag von 2009 wird explizit Bezug auf das ESFRI-Programm der EU genommen und die Initiierung einer nationalen Roadmap versprochen (vgl. o.V. 2009: 66).

14 http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri

15 <http://public.web.cern.ch/public/>

16 <http://www.awi.de/de/infrastruktur/schiffe/polarstern/>

17 http://ec.europa.eu/research/horizon2020/index_en.cfm

18 Exemplarisch sei hier auf die landesweite Arbeitsgruppe „eScience“, initiiert vom Ministerium für Wissenschaft, Forschung und Kunst in Baden-Württemberg (vgl. z.B.

sich die wissenschaftlichen Bibliotheken, vielleicht sogar in enger Kooperation mit anderen wissenschaftlich relevanten Service-Infrastruktureinrichtungen wie Rechenzentren, an verantwortlicher Stelle einbringen, eine tragende Rolle bei der weiteren Ausgestaltung und Umsetzung spielen, bleibt noch zu klären.

Wissenschaftliche Bibliotheken

Es ist nicht ganz einfach – zumindest im Internet –, eine gute Beschreibung und Definition für „wissenschaftliche Bibliotheken“ zu erhalten. Sowohl Wikipedia¹⁹ als auch das Bibliotheksportal²⁰ liefern beispielsweise nur unbefriedigende Ergebnisse. Auch beim Deutschen Bibliotheksverband – Sektion 4²¹, zuständig für wissenschaftliche Universalbibliotheken, findet sich keine Beschreibung eines Aufgabenspektrums oder eine Definition.

Das Niedersächsische Ministerium für Wissenschaft und Kultur beschreibt die Aufgaben von wissenschaftlichen Bibliotheken, zu denen die Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB) gehört, wie folgt (Stand Januar 2013):²²

„Die wissenschaftlichen Bibliotheken leisten für Forschung und Lehre die zentrale Informationsversorgung an den Wissenschaftsstandorten in Niedersachsen. [...] sind die niedersächsischen Bibliotheken nach dem Konzept der ‚Hybriden Bibliothek‘ Service-Partner der Lernenden, Lehrenden und Forscher gleichermaßen.“

Das Institut für Bibliotheks- und Informationswissenschaft (IBI) an der Humboldt-Universität zu Berlin, welches bisher das einzige universitäre Institut in Deutschland in diesem Bereich ist, beschreibt in seinem Leitbild für Forschung und Lehre den Fokus des Institutes IBI wie folgt:

DINI-Jahresprogramm 2012, <http://www.dini.de/veranstaltungen/jahrestagungen/2012/programm/>), oder auf das LOEWE-Programm des Ministeriums für Wissenschaft und Kunst in Hessen (http://www.hmwk.hessen.de/irj/HMWK_Internet?uid=fa560c0b-ed11-9311-1010-43bf5aa60dfa) hingewiesen.

19 http://de.wikipedia.org/wiki/Wissenschaftliche_Bibliothek

20 <http://www.bibliotheksportal.de/bibliotheken/bibliotheken-in-deutschland/bibliothekslandschaft/wissenschaftliche-bibliotheken.html>

21 <http://www.bibliotheksverband.de/fachgruppen/sektionen/sektion-4.html>

22 http://www.mwk.niedersachsen.de/portal/live.php?navigation_id=6353&article_id=18794&psmand=19

„[...] sind die Forschungsbestrebungen unter Wahrung traditioneller bibliothekswissenschaftlicher Themen insbesondere darauf ausgerichtet, den sich vollziehenden Paradigmenwechsel der wissenschaftlichen Kommunikation zu begleiten und der Spezifik der Digitalisierung Rechnung zu tragen. Das Zusammenwirken von Menschen und Technologien innerhalb von Informations- und Wissensprozessen und bei der Entwicklung neuer Kommunikationskulturen steht im Mittelpunkt der Forschung und Lehre am Institut.“

Interessant an beiden Aufgabenbeschreibungen ist, dass die Unterstützung für die Forschung und Lehre jeweils explizit angesprochen wird, wobei einerseits der Schwerpunkt auf die Informationsversorgung und andererseits auf die wissenschaftliche Kommunikation(skultur) gelegt wird. Beides greift in der heutigen und zukünftigen Wissens- und Informationsgesellschaft mit Stichworten wie „big data“²³, digitale Forschungsinfrastrukturen (s.u.) oder Forschungsdateninfrastrukturen (s.u.) noch zu kurz.

Die Abteilung Forschung & Entwicklung der SUB Göttingen, die sich v.a. auf die Unterstützung des digitalen Forschungsprozesses unterschiedlichster Fachdisziplinen konzentriert, beschreibt das Aufgabenprofil wie folgt:²⁴

„Wir versuchen aufkommende Entwicklungen der Informations- und Wissensgesellschaft frühzeitig zu erkennen, aufzugreifen und für die Wissenschaft nutzbar zu gestalten – und dadurch konstruktiv an einer Neupositionierung der Rolle von wissenschaftlichen Bibliotheken im 21. Jahrhundert mitzuwirken. [...] stehen im Dienste der Wissenschaft und decken inhaltlich jeweils unterschiedliche Bereiche in der Beratung und Konzeptionierung bis hin zur prototypischen Software-Entwicklung ab. [...] Wir wollen damit den Grundstein dafür legen, Forschungsergebnisse langfristig interpretierbar zu machen und durch Vernetzung sowie Abstraktion (z.B. Visualisierung) in neuen Zusammenhängen kontextualisieren zu können.“

Es wäre wünschenswert, in Deutschland gemeinsam über eine Neuorientierung bzw. Erweiterung des Aufgaben- und Kompetenzprofils von wissenschaftlichen Bibliotheken nachzudenken, die u.a. sowohl analoge Angebote und digitale Informationsversorgung als auch wissenschaftliche Kommunikationskulturen sowie die Unterstützung und Mitgestaltung von Forschungsprozessen aller Wissenschaftsgebiete umfasst. Sicherlich lässt sich diese Aufzählung noch erweitern und muss von Zeit zu Zeit aktualisiert werden, um mit dem rasanten Technologiewechsel und einhergehenden Veränderungs-

23 Die Diskussion um „big data“ ist bereits im Alltag angekommen, vgl. z.B. *Die Zeit online* mit einem entsprechendem Artikel (vgl. Fischermann/Hamann 2013).

24 <http://www.rdd.sub.uni-goettingen.de/>

prozessen in den (digitalen) Forschungsprozessen der verschiedenen Fachdisziplinen Schritt halten zu können. Denn über eines besteht Konsens: Die wissenschaftlichen Bibliotheken sind Service- und Infrastrukturpartner der Wissenschaft. Vielleicht ist es aber auch an der Zeit, darüber nachzudenken, den wissenschaftlichen Bibliotheken eigene Forschungs- und Entwicklungsaufgaben zuzuweisen, wie es an zahlreichen wissenschaftlichen Rechenzentren zum Beispiel bereits Usus ist.

Forschungsinfrastrukturen

Forschungsinfrastruktur ist ein Begriff, den es auch vor dem digitalen Zeitalter gab. Er wurde und wird genutzt, um die Konzeption, die Entwicklung und den Betrieb großer Infrastrukturen zu gewährleisten, um bestimmte, aktuelle und relevante Forschungsfragen durch geeignete organisatorische, technologische, rechtliche und finanzielle Rahmenbedingungen zu unterstützen. Das Forschungsschiff Polarstern zum Beispiel hat „seit ihrer Indienststellung am 9. Dezember 1982 [...] über fünfzig Expeditionen in Arktis und Antarktis abgeschlossen. Sie wurde eigens für die Arbeit in den Polarmeeren konzipiert und ist gegenwärtig eines der leistungsfähigsten Polarforschungsschiffe der Welt. [...] Das Schiff ist für biologische, geologische, geophysikalische, glaziologische, chemische, ozeanografische und meteorologische Forschungsarbeiten ausgerüstet und verfügt über neun wissenschaftliche Labors“²⁵. Es gibt zahlreiche weitere Beispiele für analoge Forschungsinfrastrukturen und seit einigen Jahrzehnten insbesondere in den Natur- und Lebenswissenschaften parallel Beispiele für digitale Forschungsinfrastrukturen.²⁶ In den Geistes- und Sozialwissenschaften gibt es erst recht wenige digitale Forschungsinfrastrukturen und auch erst seit ca. zehn Jahren, als im sog. EU-ESFRI²⁷-Programm zum ersten Mal die sog. SSH²⁸-Fächer durch die Europäische Kommission aufgefordert wurden, ebenfalls digitale Forschungsinfrastrukturen im Rahmen der sog. „preparatory phase“ (Vorbereitungs- bzw. Konzeptionsphase) zu entwickeln. Mittlerweile gibt es fünf digi-

25 <http://www.awi.de/de/infrastruktur/schiffe/polarstern/>

26 In der aktuellen ESFRI Roadmap sind alle derzeit laufenden Vorhaben gelistet, s. http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri-roadmap.

27 http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri

28 Social Science and Humanities

tale Forschungsinfrastrukturen in den Geistes- und Sozialwissenschaften, die bereits in die 2. Phase, die Entwicklungsphase, eingetreten sind und sich mithilfe eines ERIC²⁹ (European Research Infrastructure Consortium) zu einer Rechtsform nach europäischem Recht zusammengeschlossen haben bzw. kurz davor stehen:

- CESSDA – Council of European Social Science Data Archives³⁰
- CLARIN – Common Language Resources and Technology Infrastructure³¹
- DARIAH – Digital Research Infrastructure for the Arts and Humanities³²
- ESS – European Social Survey³³
- SHARE – Survey of Health, Ageing and Retirement in Europe³⁴

Im Folgenden wird die DARIAH-Forschungsinfrastruktur aus deutscher Perspektive näher vorgestellt.

DARIAH-DE, der deutsche Beitrag zu DARIAH-EU, wird vom BMBF mit einer Laufzeit von insgesamt fünf Jahren für die Entwicklungsphase (im EU-ESFRI-Kontext die sog. construction phase) finanziert, weitere EU-Länder beteiligen sich ebenfalls an den Entwicklungskosten und tragen zur europäischen Forschungsinfrastruktur mit speziellen Diensten und Angeboten bei. DARIAH-DE steht vor der Herausforderung, sowohl eine eigenständige Forschungsinfrastruktur aufzubauen, die nationale Forschungsprojekte unterstützt, als auch komplementär zu einer digitalen Gesamt-Forschungsinfrastruktur auf europäischem Level beizutragen. Die beteiligten EU-Partner haben sich darauf verständigt, dass mindestens im DARIAH-ERIC-Verbund die Dienste und Daten so weit als möglich kostenfrei genutzt werden können, sodass Werkzeuge, die z. B. in einem EU-Partnerland entwickelt werden, auch in Deutschland und für deutsche Forschungsprojekte genutzt werden können. Die DARIAH-DE-Forschungsinfrastruktur ist eine eigenständige, in sich abgerundete Entwicklung, die sich allerdings sinnvoll in den „Masterplan“ von DARIAH-EU einfügt.

29 http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=eric

30 <http://www.CESSDA.org/>

31 <http://www.clarin.eu/>

32 <http://www.dariah.eu/>

33 <http://www.europeansocialsurvey.org/>

34 <http://www.share-project.org/>

Ziel von DARIAH ist es, die in den Geistes- und Kulturwissenschaften mit digitalen Methoden und Hilfsmitteln durchgeführte Forschung zu unterstützen und weiter voranzutreiben, verschiedene geistes- und kulturwissenschaftliche Fachdisziplinen miteinander zu vernetzen und durch den Aufbau einer digitalen Forschungsinfrastruktur den Austausch von Ressourcen, Methoden, Daten und Erfahrungen zu fördern.

Das Überschreiten des Horizonts der traditionellen Fragen in den Geistes- und Kulturwissenschaften ist durch die seit gut einem Jahrzehnt in immer größerer Zahl verfügbaren digitalen Repräsentationen kultureller Äußerungen und Artefakte möglich.

Mit DARIAH verbindet sich die Vision, den Geistes- und Kulturwissenschaften in Europa und in Deutschland ebenfalls einen Raum zur Großforschung zu schaffen, wie er in den Natur- und angewandten Wissenschaften bereits seit Jahrzehnten von vielen Forschern und Institutionen gemeinsam aufgebaut und genutzt wird. Dadurch ergibt sich die einmalige Chance, die Geistes- und Kulturwissenschaften an Spitzentechnologie teilhaben zu lassen und ihnen neue Wege und Möglichkeiten der Zusammenarbeit und des Forschens zu eröffnen.

Forschungsinfrastrukturen sind kein statisches Produkt, sondern sind als kontinuierlicher Prozess zu betrachten, der von der Forschung angetrieben werden muss. Dieser Prozess beinhaltet auch eine starke Vernetzung einerseits der Fachwissenschaftler untereinander und andererseits mit IT-Experten, Informationswissenschaftlern und Informatikern.

Ausgehend von dem sich zurzeit weiterhin vollziehenden Paradigmenwechsel auch in den Geisteswissenschaften, besteht das Ziel von DARIAH-DE darin, durch die Entwicklung einer modernen Forschungsinfrastruktur für die Geistes- und Kulturwissenschaften einen Beitrag für deren Wettbewerbs- und Anschlussfähigkeit im internationalen Forschungskontext zu leisten und den europäischen Forschungsraum und damit DARIAH-EU aktiv mitzugestalten. Einige Forschungsfragen lassen sich mit traditionellen Methoden nach wie vor sehr gut beantworten, andere, die sich aus den neu verfügbaren Forschungsobjekten und daraus entstandenen Daten erst ergeben, bedürfen neuer Methoden und Instrumente für ihre Analyse und für weiterführende Diskussionen.

DARIAH-DE arbeitet konkret entlang folgender Schwerpunktthemen daran, den neuen Entwicklungen und fachwissenschaftlichen Anforderungen Rechnung zu tragen und insgesamt die Digital Humanities weiter zu befördern:

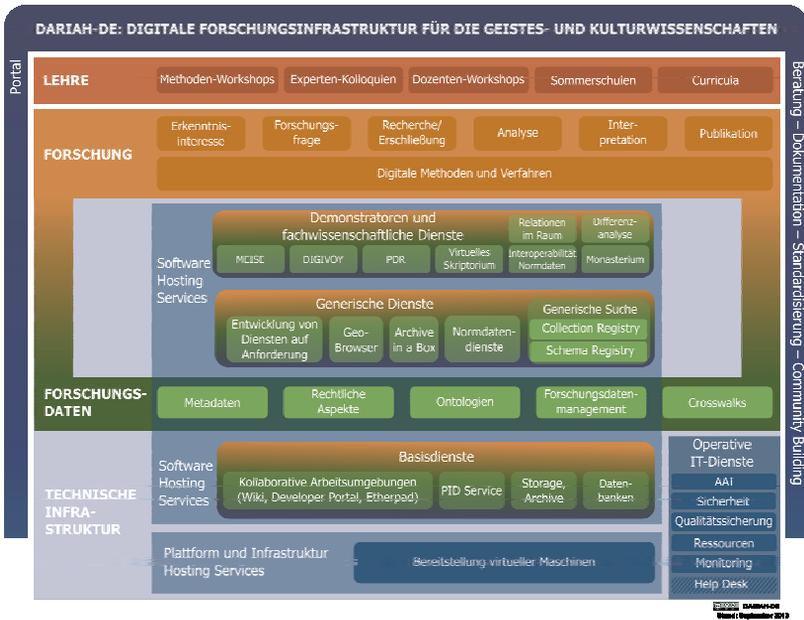


Abb. 1 Übersicht über die DARIAH-DE-Forschungsinfrastruktur (Stand Sept. 2013)

- **Lehre:** Nachwuchsförderung und Qualifizierungsmaßnahmen, z.B.
 - Entwicklung von Konzepten für und Durchführung von Curricula, Expertenseminaren, Summer Schools, Dozenten- und Methodenworkshops etc.
 - Community-Building
- **Forschung:** Forschungszyklus (Lifecycle), fachwissenschaftliche Dienste, z.B.
 - Entwicklung, Kartierung und Dokumentation von Methoden und Verfahren speziell für die digitalen Geistes- und Kulturwissenschaften
- **Forschungsdaten:** wissenschaftliche Sammlungen, generische Dienste, Forschungsdatenmanagement, z.B.
 - Entwicklung von Konzepten zum standardisierten Datenaustausch (Crosswalks)
 - Empfehlungen, Best Practices, Guidelines z.B. zum Forschungsdatenmanagement, zu Normdatendiensten, fachspezifischen Metadaten
 - Entwicklung der Collection- und Schema-Registry

- *technische Infrastruktur*: generische Basisdienste, z.B.
 - Bereitstellung kollaborativer Arbeitsumgebungen wie z.B. Etherpad, Developer Portal etc.
 - Monitoring der DARIAH-DE und externen Diensten aus affilierten Projekten
 - Authentifizierungsdienste, die auch auf EU-Ebene eingesetzt werden
 - persistente Identifizierungsdienste für die Zitier- und Publikationsfähigkeit von Artikeln und Forschungsdaten/-sammlungen

Die Aspekte „Lehre und Forschung“ werden auf DARIAH-EU-Ebene durch das Virtuelle Kompetenzzentrum VCC 2 „Research & Education“ vertreten, Forschungsdaten durch das Virtuelle Kompetenzzentrum VCC 3 „Research Data“ und die technische Infrastruktur durch das Virtuelle Kompetenzzentrum VCC 1 „eInfrastructure“.

DARIAH-DE orientiert sich an aktuellen konkreten Forschungsfragen aus unterschiedlichen Fachwissenschaften (z.B. Sprach- und Literaturwissenschaften, Musikwissenschaften, Archäologie, Judaistik etc.) und konzipiert entlang dieser Forschungsfragen, der eingesetzten digitalen Methoden und der zugrunde liegenden wissenschaftlichen Sammlungen (Forschungsdaten) fachwissenschaftliche Dienste. Diese sogenannten Demonstratoren erfüllen in der Regel nicht nur Anforderungen aus einer bestimmten Fachdisziplin, sondern unterstützen übergreifend auch Bedürfnisse aus anderen Fachwissenschaften (z. B. Geobrowser zur Visualisierung in Raum und Zeit). Damit versteht sich DARIAH-DE als forschungsgetrieben, d. h., es werden keine „in Stein gemeißelten Beton-Silos“ aufgebaut, sondern aktuelle Forschungsfragen mit ihren verschiedenen und zum Teil gleichen digitalen Methoden und Verfahren bestimmen die Ausrichtung der Forschungsinfrastruktur und treiben deren Konzeption und Entwicklung an.

Mehrere externe fachwissenschaftliche Dienste sind so konzipiert, dass sie in das DARIAH-Framework übernommen und somit nach Ablauf der jeweiligen Förderphase weiter angeboten werden können. Zahlreiche europäisch (z.B. die EU-Projekte EHRI³⁵, Cendari³⁶) und deutsch (z.B. RiR³⁷, Mo-

35 European Holocaust Research Infrastructure, <http://www.ehri-project.eu/>

36 Collaborative European Digital Archive Infrastructure, <http://www.cendari.eu/>

37 Relationen im Raum. Visualisierung topografischer Klein(st)strukturen, <http://www.steinheim-institut.de/wiki/index.php/RiR>

nasterium.Net³⁸) finanzierte Forschungsprojekte greifen jetzt schon auf bestimmte Dienste von DARIAH-DE zurück. Auch damit ist sichergestellt, dass DARIAH-DE aktuelle Forschungsanforderungen aufnimmt und diese als Basis für die weitere Entwicklung dienen. Zurzeit greifen ca. 400 Fachwissenschaftler unterschiedlichster Disziplinen aus verschiedenen EU-Ländern auf einzelne Dienste in DARIAH-DE zu. Diese Zahl dürfte bei Fertigstellung weiterer (fachwissenschaftlicher) Dienste und dem Ausbau des digitalen infrastrukturellen Angebots zu Februar 2014 stark steigen.

Neue Rolle und Aufgaben für wissenschaftliche Bibliotheken

Die wissenschaftliche Bibliothek ist als zentrale Service-Infrastruktureinrichtung an einer Universität ein kostenintensiver Faktor, mancherorts dürften die Kosten auf einem Universitätscampus für die Bibliothek die höchsten sein. Mit ihrer zentralen Aufgabe, die Forschung und Lehre am Standort und darüber hinaus zu unterstützen, müssen sich wissenschaftliche Bibliotheken auch mit den Veränderungen in den Forschungsprozessen auseinandersetzen, wenn sie ihre Rolle als Kooperationspartner der Wissenschaftsdisziplinen nicht verlieren wollen. Traditionelle Aufgaben im Zeitalter der kooperativen Verbundkatalogisierung zum Beispiel können (weiter) eingeschränkt werden zugunsten von Angeboten im Bereich elektronisches Publizieren, Open Access und Universitätsverlage. Aber wissenschaftliche Bibliotheken müssen sich wesentlich mehr engagieren und die fachwissenschaftlichen Disziplinen in Zukunft besser und gezielter unterstützen. Dazu gehört auch, dass die Forschungsprozesse und die dort eingesetzten neuen digitalen Methoden und Verfahren besser verstanden und in Kooperation mit Fachwissenschaftlern gemeinsam entwickelt und bereitgestellt werden. Hier kommen auf wissenschaftliche Bibliotheken neue Herausforderungen zu, die aber als konsequente Fortschreibung ihrer bisherigen Kompetenz und Erfahrung verstanden werden können. Bibliotheken waren schon immer eine erste Anlaufadresse, um Information und Wissen zu organisieren, aufzubereiten und für die Nutzung zur Verfügung zu stellen. Dies trifft weiterhin für analoge Angebote zu und wird mehr und mehr ergänzt durch digitale Dienste. Dies bedeutet zum Beispiel, dass sich wissenschaftliche Bibliotheken auch mit dem „digitalen Datenbereich“ auseinandersetzen müssen. Allgemein kann festgehalten

38 <http://www.mom-ca.uni-koeln.de/mom/home>

werden, dass sich wissenschaftliche Bibliotheken mit folgenden Aspekten beschäftigen:

- (fachwissenschaftliche) Informations- und Datenaufbereitung zur Strukturierung und Beschreibung von Informationen und Daten
- Bereitstellung von (fachwissenschaftlichen) Informationsdiensten und Werkzeugen, um Informationen und Daten zu verarbeiten und überhaupt nutzbar zu machen
- Verständnis über Informationsmanagement und über fachwissenschaftliche Forschungsprozesse, um Daten und Werkzeuge sinnvoll im Forschungskontext und seinen Prozessen anzuwenden
- Kenntnisse von (innovativen) Informationstechnologien und angewandten Informatikkomponenten für die Konzeption und Entwicklung von digitalen Diensten

Konkret auf aktuelle Fragestellungen konzentriert, denen sich Experten aus den wissenschaftlichen Bibliotheken gemeinsam mit Experten aus den Rechenzentren und den Fachdisziplinen widmen können, bedeutet es, sich mit diesen Themenfeldern auseinanderzusetzen (diese Liste ist sicherlich nicht vollständig und muss weiter ausgearbeitet werden):

- *Forschungsdaten, Forschungsdatenmanagement, Langzeitarchivierung:*
 - Wann und wie entstehen digitale Datensammlungen? Welche Datentypen fallen an und welche Datenformate müssen auch über die Zeit aufbewahrt werden?
 - Wie können diese standardisiert beschrieben werden (bibliografische Daten, Kontext-Daten, fachspezifische Anreicherung)?
 - Wie können diese Daten für den Zeitraum des Forschungsprojektes gesammelt und aufbewahrt werden?
 - Welche Daten müssen auch nach Ablauf eines Forschungsprojektes gespeichert und zugänglich gemacht werden und in welcher Form (z.B. redundante Speicherung, Nachweis in einem fachspezifischen Datenzentrum etc.)?
 - Welche Richtlinien und Daten-Policy gibt sich ein Forschungsverbund, v.a. in internationalen Kooperationsverbänden?
- *kooperative und kollaborative Arbeitsumgebungen, Virtuelle Forschungsumgebungen:*
 - Wie kann das kooperative Arbeiten der Wissenschaftler in einem Forschungsverbund unterstützt werden (z.B. Wiki, gemeinsames Code- oder Dokumenten-Repository)?

- Welche Werkzeuge braucht es dafür? Welche bereits existierenden Werkzeuge können nachgenutzt werden (z.B. für die Annotation von Text)?
- Welche Rechte- und Rollenmodelle braucht ein Forschungsverbund, um z.B. festzuhalten, wer wann wie auf bestimmte Datensätze zugreifen darf?
- Wie sieht ein Publikationsworkflow aus und wie kann er umgesetzt werden?
- *technische Infrastruktur*
 - Welche technischen Basisstrukturen müssen zur Verfügung gestellt werden (z. B. Datenbanken, Speicherplatz, Rechenleistung, virtuelle Maschinen etc.)?
 - Welche generischen Basisdienste werden benötigt (z. B. persistente Identifizierungen für Daten und Publikationen, Authentifizierungs- und Autorisierungsmechanismen etc.)?
 - Wie können fachwissenschaftliche Werkzeuge (bereits extern vorhandene oder eigens entwickelte) eingebunden werden in die Forschungsumgebung?
- *Schulungs- und Lehrangebote im Bereich Informationswissenschaften:*
 - Was bedeutet wissenschaftliche gute Praxis und was kann jeder einzelne Wissenschaftler in seinem Forschungsverbund dazu beitragen?
 - Wie werden Datensammlungen in einer Art und Weise beschrieben, dass auch Jahre später dokumentiert ist, wie die Daten entstanden sind und wie sie nachgenutzt werden können?
 - Wie können fachspezifische Datensammlungen für die Suche und Nachnutzung bereitgestellt werden?
 - Wie können Datensammlungen publiziert werden und was wird dafür alles benötigt?

Diese Aufzählung ist nicht vollständig und gibt nur einen kleinen Einblick in das neue Aufgabenspektrum, in das sich wissenschaftliche Bibliotheken einbringen können. Ein Blick in den anglo-amerikanischen Raum zeigt, dass aktuelle Stellenausschreibungen der dortigen Bibliotheken mehr und mehr ein Stellenprofil an der Schnittstelle Bibliothek/Fachwissenschaften veröffentlichen:

- *Data Management Consultant* at the Johns Hopkins University Sheridan Libraries (veröffentlicht im Oktober 2012 über die DLF³⁹-E-Mail-Liste)
- Brown University Library: *Scientific Data Curator* (veröffentlicht im Oktober 2012 über die DLF-E-Mail-Liste)
- *Research Data Manager*, reporting to the Director of the Center for Digital Research and Scholarship (CDRS), a division of the Columbia University Libraries/Information Services (veröffentlicht im September 2012 über die DLF-E-Mail-Liste)

Diese Liste ließe sich beliebig fortsetzen und eine systematische Auswertung der aktuellen Stellenbeschreibungen an anglo-amerikanischen Universitätsbibliotheken wäre sehr interessant. Der Trend, dass mehr und mehr bibliotheksrelevante Stellenprofile mit Daten, ihrer Aufbereitung, ihrem Management und Veröffentlichung zu tun haben, ist klar erkennbar. In Deutschland gibt es sehr wenige Hochschulen und wohl nur eine Universität (HU Berlin), die im Bereich Informationswissenschaft ausbildet und den einen oder anderen neuen Themenschwerpunkt in ihren Curricula verankert. Allerdings gibt es bisher keinen Studiengang, der zum Beispiel zu einem Profil „data librarian“ oder „data scientist“ führt. Hierfür wäre es auch notwendig, dass die „klassische“ Informationswissenschaft sehr eng mit der Informatik kooperiert, wie dies bspw. die Humboldt-Universität zu Berlin am *Institut für Bibliotheks- und Informationswissenschaft* in Kooperation mit dem *Institut für Informatik* umsetzt. Ergebnis dieser Kooperation ist die Möglichkeit, einen Monobachelor Informationsmanagement & Informationstechnologie⁴⁰ zu studieren.

Dazu passt auch ein Artikel, der im Oktober 2012 im *Harvard Business Review* erschienen ist: „Data Scientist: The Sexiest Job of the 21st Century“ von Thomas H. Davenport und D. J. Patil (Davenport/Patil 2012). Dieser Artikel zeigt eindrucksvoll, welches Potenzial im Bereich (großer) digitaler Datenmengen und ihrer Auswertung und Visualisierung noch liegt und welche spannenden Aufgaben noch vor uns liegen, auch in Bereichen, in denen wissenschaftliche Bibliotheken nicht immer involviert sein müssen. Um mit digitalen großen (z.B. Lebens- und Naturwissenschaften) oder heterogenen (z.B. Geistes- und Kulturwissenschaften) Daten umgehen zu können, braucht

39 Vgl. <http://lists.clir.org/scripts/wa-CLIR.exe?INDEX>, die einzelnen Stellenausschreibungen sind nicht mehr zugänglich.

40 <http://www.ibi.hu-berlin.de/studium/infomit/>

es stabile digitale Forschungsinfrastrukturen, bei deren Konzeption, Entwicklung und Bereitstellung wissenschaftliche Bibliotheken mitwirken können und sollten.

Zusammenfassung und Ausblick

(Wissenschaftliche) Bibliotheken bieten seit Jahrtausenden eine Forschungsinfrastruktur für Wissenschaftler und Wissbegierige an. Während in früheren Zeiten dazu der Interessierte zu der Bibliothek kommen und sich vor Ort mit den Publikationen beschäftigen musste, ist es seit einigen Jahrzehnten möglich, sich digitale Objekte zu jeder Zeit von jedem Ort aus anzuschauen, auszuleihen und nachzunutzen. Der rasche Zuwachs an digitalen Objekten – und hier nicht nur Publikationen, sondern auch zunehmend Forschungsdaten – erlaubt eine wesentlich größere Flexibilität im wissenschaftlichen Erkenntnisprozess. Mehrere Wissenschaftler aus verschiedenen Orten/Ländern können sich gemeinsam mit der Forschungsfrage beschäftigen, dazu relevante digitale Daten nachnutzen sowie digitale Methoden und Verfahren anwenden, soweit geeignete Werkzeuge und Services zur Verfügung stehen. In diesem Sinne nutzen sie auch eine Forschungsinfrastruktur, die angepasst an die heutigen Verhältnisse in digitaler Form vorliegt. Die entscheidende Frage wird sein, welche Rolle wissenschaftliche Bibliotheken dabei noch einnehmen (können), ob sie sich zum Beispiel aktiv an der Konzeption und Entwicklung von (fachspezifischen) digitalen Forschungsinfrastrukturen beteiligen oder ob diese Rolle andere Stakeholder (z. B. Rechenzentren, fachspezifische Datenzentrums-Betreiber etc.) einnehmen werden. Insgesamt werden die nächsten Jahre zeigen, welche Aufgabengebiete und Verantwortlichkeiten bei wissenschaftlichen Bibliotheken verbleiben oder ob ihre Bedeutung angesichts des rasanten Technologieumzugs und -wechsels auf digitale Infrastrukturen abnehmen wird. Die Erfahrungen der letzten zehn Jahre in der SUB-Abteilung Forschung & Entwicklung haben gezeigt, dass eine Kooperation mit verschiedenen Fachdisziplinen auf Augenhöhe möglich ist, um die „digitale Transformation“⁴¹ gemeinsam voranzutreiben. Eine große Herausforderung besteht sicherlich darin, die wissenschaftlichen Bibliotheken so umzubauen, dass sie den neuen Aufgabengebieten in Zukunft gewachsen sind. Dies setzt eine transparente Personalstrategie, eine konsequente Personalent-

41 Vgl. Positionspapier der DFG dazu (DFG 2012a).

wicklung, einen Gesamtfinanzierungsplan über mehrere Jahre, eine exzellente Kommunikation in die Bibliothek, auf dem Campus und nach außen zu den Fachdisziplinen sowie generell ein „Strategie für die digitale Transformation“ voraus, die idealerweise von den Führungskräften der Bibliothek mitgetragen wird.

Literaturverzeichnis

- Allianzinitiative (2013): Allianz der Wissenschaftsorganisationen setzt die Schwerpunktinitiative „Digitale Information“ fort. http://www.allianzinitiative.de/de/steuerungsgremium/allianz_der_wissenschaftsorganisationen_setzt_die_schwerpunktinitiative_digitale_information_fort/.
- Davenport, T. H.; Patil, D. J. (2012): Data Scientist: The Sexiest Job of the 21st Century. <http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/ar/1#>.
- Deutscher Bundestag – Enquete-Kommission Internet und digitale Gesellschaft (2011): Projektgruppensitzung als öffentliches Expertengespräch am 7. November 2011. http://www.bundestag.de/internetenquete/dokumentation/Bildung_und_Forschung/PGBuF_2011-11-07_Expertengespraech/index.jsp.
- Deutscher Bundestag – Enquete-Kommission Internet und digitale Gesellschaft (2012): Projektgruppe Bildung und Forschung – Handlungsempfehlungen. http://www.bundestag.de/internetenquete/dokumentation/Sitzungen/20120625/A-Drs_17_24_052_-_PG_Bildung_und_Forschung_Handlungsempfehlungen.pdf.
- DFG – Deutsche Forschungsgemeinschaft (2010): Informationsverarbeitung an Hochschulen – Organisation, Dienste und Systeme. http://www.dfg.de/download/pdf/foerderung/programme/wgi/empfehlungen_kfr_2011_2015.pdf.
- DFG – Deutsche Forschungsgemeinschaft (2012a): Die digitale Transformation weiter gestalten – Der Beitrag der Deutschen Forschungsgemeinschaft zu einer innovativen Informationsinfrastruktur für die Forschung. http://www.dfg.de/download/pdf/foerderung/programme/lis/positionspapier_digitale_transformation.pdf.
- DFG – Deutsche Forschungsgemeinschaft (2012b): Stellungnahme zu den Empfehlungen des Wissenschaftsrates zur „Weiterentwicklung der Informationsinfrastrukturen in Deutschland bis 2020“. http://www.dfg.de/download/pdf/foerderung/programme/lis/stellungnahme_dfg_informationsinfrastrukturen.pdf.
- EU – High Level Expert Group on Scientific Data (2010): Riding the wave – How Europe can gain from the rising tide of scientific data. <http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf>.

- Fischermann, Thomas; Hamann, Götz (2013): Wer hebt das Datengold? In: *Die Zeit online* vom 6. Januar 2013. <http://www.zeit.de/2013/02/Big-Data>.
- HRK – Hochschulrektorenkonferenz (2012): Hochschule im digitalen Zeitalter: Informationskompetenz neu begreifen – Prozesse anders steuern. http://www.hrk.de/uploads/media/Entschliessung_Informationskompetenz_2012012.pdf.
- Neuroth, Heike (2012): Aktuelle und künftige Forschungsaufgaben. In: Konrad Umlauf, Stefan Gradmann (Hrsg.): *Handbuch Bibliothek: Geschichte, Aufgaben, Perspektiven*. Stuttgart u.a.: Metzler, S. 218–227.
- o.V. (2009): Wachstum. Bildung. Zusammenhalt. Koalitionsvertrag zwischen CDU, CSU und FDP. http://www.bmi.bund.de/SharedDocs/Downloads/DE/Ministerium/koalitionsvertrag.pdf?__blob=publicationFile.
- Wissenschaftsrat (2011): Übergreifende Empfehlungen zu Informationsinfrastrukturen. <http://wisspub.net/2011/01/31/informationsinfrastruktur-wissenschaftsrat/>.
- Wissenschaftsrat (2012): Empfehlungen zur Weiterentwicklung der wissenschaftlichen Informationsinfrastrukturen in Deutschland bis 2020. <http://www.wissenschaftsrat.de/download/archiv/2359-12.pdf>.

Aus Sicht der Geisteswissenschaften: Die wissenschaftliche Bibliothek als Schnittstelle zwischen digitaler Infrastruktur und geisteswissenschaftlicher Forschung

von Andrea Rapp

Die Gelegenheit, der Abteilung Forschung & Entwicklung an der SUB Göttingen zum zehnjährigen Bestehen zu gratulieren, will ich mir nicht nehmen lassen, namentlich Prof. em. Dr. Dr. h.c. mult. Elmar Mittler, Prof. Dr. Norbert Lossau und Dr. Heike Neuroth, deren Weitblick und Tatkraft es zu verdanken ist, dass diese Abteilung und ihr großes Team wichtige Kooperationspartner für Forschende in den Geisteswissenschaften sind – Glückwunsch, Respekt und Dankbarkeit für diese Leistung! Das in dieser Kooperation liegende Potenzial können das Göttinger Festkolloquium und dieser Band nun einmal mehr eindrucksvoll bezeugen.

Die Anforderung, aus der Sicht der Geisteswissenschaften einen Ausblick auf die Evolution der Forschungsinfrastruktur in der Kooperation von Bibliothek und Forschung zu geben, kann kaum angemessen erfüllt werden, nimmt man allein die – auch in diesem Band häufig angesprochene – Vielfalt und Heterogenität der „Geisteswissenschaften“ in den Blick, die im Übrigen gewiss kein Nachteil, aber eben doch eine Herausforderung ist (vgl. Wissenschaftsrat 2005; BMBF 2007). Die folgenden kurzen Überlegungen können daher nur aus einer begrenzten Perspektive heraus erfolgen und sind durch meine persönlichen Erfahrungen geprägt. Zudem beschränke ich mich auf die Teile der Geisteswissenschaften, die sich selbst digital verorten. Der Verortungs- und Definitionsprozess der „Digital Humanities“¹ ist derzeit im Gange, wobei das Spektrum der digitalen Affiliierung ebenso heterogen und vielfältig ist wie die Geisteswissenschaften selbst: Von der Nutzung digital bereitgestellter Quellen über die Erstellung digitaler Materialien, Editionen und Sammlungen sowie digitaler Publikationen bis hin zur Nutzung, Erforschung und Entwicklung digitaler Methoden sind die Übergänge fließend.

¹ Ein breiter Nachweis der Debatte ist an dieser Stelle nicht möglich; exemplarisch herausgegriffen sei Thaller (2012).

Die Geisteswissenschaften haben die geistigen, künstlerischen und kulturellen Hervorbringungen des Menschen zum Gegenstand, die u.a. auch mit historischen, politischen, soziologischen und religiösen Aspekten verbunden sind. In Diltheyscher Tradition können die Geisteswissenschaften plakativ verkürzt als Wissenschaften vom Verstehen gegen die Naturwissenschaften als Wissenschaften vom Erklären abgegrenzt werden. In die geisteswissenschaftliche Tradition lassen sich auch die Digital Humanities einschreiben (Humanities Computing ist die ältere Bezeichnung):

“Humanities computing is precisely the automation of every possible analysis of human expression (therefore, it is exquisitely a ‘humanistic’ activity), in the widest sense of the word, from music to the theater, from design and painting to phonetics, but whose nucleus remains the discourse of written texts.” (Busa 2004)

Dennoch reklamieren die Digital Humanities durch den Einsatz computergestützter Verfahren eine entscheidende Erweiterung und Erneuerung der Geisteswissenschaften:

“What this collection also reveals is that there are central concerns among digital humanists which cross disciplinary boundaries. This is nowhere more evident than in the representation of knowledge-bearing artifacts. The process of such representation – especially so when done with the attention to detail and the consistency demanded by the computing environment – requires humanists to make explicit what they know about their material and to understand the ways in which that material exceeds or escapes representation. Ultimately, in computer-assisted analysis of large amounts of material that has been encoded and processed according to a rigorous, well thought-out system of knowledge representation, one is afforded opportunities for perceiving and analyzing patterns, conjunctions, connections, and absences that a human being, unaided by the computer, would not be likely to find.

The process that one goes through in order to develop, apply, and compute these knowledge representations is unlike anything that humanities scholars, outside of philosophy, have ever been required to do. This method, or perhaps we should call it a heuristic, discovers a new horizon for humanities scholarship, a paradigm as powerful as any that has arisen in any humanities discipline in the past – and, indeed, maybe more powerful, because the rigor it requires will bring to our attention undocumented features of our own ideation. Coupled with enormous storage capacity and computational power, this heuristic presents us with patterns and connections in the human record that we would never otherwise have found or examined.” (Schreibman/Siemens/Unsworth 2004)

Die Digital Humanities versprechen also nicht weniger als ein neues Forschungsparadigma. Der *digital turn* in den Geisteswissenschaften soll die

geisteswissenschaftlichen Forschungskulturen verändern, einen Empirie-schub (*empirical turn*) und damit eine „Verwissenschaftlichung“ bewirken (vgl. Gippert 2010), alte Forschungsfragen effizient, nachvollziehbar und wiederholbar angehen sowie neue Forschungsfragen generieren und Lösungswege und Antwortmöglichkeiten in Angriff nehmen. Entscheidend hierfür sind aus meiner Sicht tatsächlich die Explizierung von implizitem Wissen, die Möglichkeiten des Zugriffs auf Massendaten bzw. größere Quellen- und Datenumfänge als zuvor und die „Strenge“ (rigor) des Arbeitens, die der Computereinsatz uns abverlangt. Begreift man die Geisteswissenschaften nach wie vor als Wissenschaften vom Diskurs und vom Verstehen, wird deutlich, dass die Verbreiterung der uns zugänglichen Datengrundlage und die Explizierung des in unseren Daten- und Wissensspeichern inhärenten Wissens genau dies befördern können, dass die Geisteswissenschaften durch die Erschließung digitaler Methoden – verbunden mit einer gewissen Annäherung an naturwissenschaftliche Zugänge – für ihre Forschungsfragen nur gewinnen können, keinesfalls etwas aufgeben müssen (vgl. Fish 2012).

Wie gehen nun die Evolution der Forschungsinfrastruktur (hier bezogen auf die Bibliothek) und die Geisteswissenschaften zusammen?

Infrastruktur – auch Forschungsinfrastruktur in einem umfassenden Sinne – bietet als Fundament zunächst den einfachen und offenen Zugang zu den Forschungsgegenständen, eine „einfache“ Tatsache, die jedoch nicht hoch genug bewertet werden kann. Die digitale Bereitstellung der Quellen und Artefakte, der geisteswissenschaftlichen Forschungsobjekte in digitaler Form, eröffnet sowohl einen leichteren Zugriff auf „mehr Quellen“ als zuvor als auch einen neuen, frischen Blick darauf. Der Transformationsprozess in ein neues Medium fordert neue Modell- und Theoriebildungen zu den digitalen Repräsentationen des kulturellen Erbes und bewirkt zugleich eine erneute Hinwendung zum Original, sodass die Forschung zu beiden medialen Erscheinungsformen neue Erkenntnisse hervorbringt. Sichtbar wird dieser „Trend“ der Hinwendung zum Original und zur Tätigkeit des Sammelns – um nur einiges exemplarisch herauszugreifen – z.B. in den aktuellen Sonderforschungsbereichen SFB 933 in Heidelberg zur Materialität,² dem SFB 950

2 Heidelberger Sonderforschungsbereich 933: Materiale Textkulturen. Materialität und Präsenz des Geschriebenen in non-typografischen Gesellschaften (<http://www.materiale-textkulturen.de/>)

zu Manuskriptkulturen in Hamburg,³ in den aktuellen Ausschreibungen des Bundesministerium für Bildung und Forschung zur Sprache der Objekte,⁴ dem DFG-Netzwerk Sprachen des Sammelns an der HU Berlin,⁵ aber auch in der Etablierung und Weiterentwicklung spezifischer Standards, z.B. entsprechender Tagsets und Special Interest Groups im Rahmen der Text-Encoding-Initiative (TEI).⁶ Die Universität Göttingen macht – über ihre herausragende Bibliothek hinaus – verstärkt wieder auf ihre universitären Sammlungen und deren Bedeutung für die Forschung aufmerksam;⁷ analoge und digitale Präsentationen ergänzen sich hier. Man kann also durchaus konstatieren, dass die Rolle der Bibliothek als Forschung ermöglichende Infrastruktureinrichtung gerade auch durch die Digitalisierung ihres Bestandes nochmals besonders gestärkt wird.⁸

Die Transition des kulturellen Erbes ins digitale Medium befördert demnach auch unsere Kenntnis der analogen Artefakte, da zum einen die Digitalisierung Implizites durch Metadaten und Annotationen explizit macht, was häufig eine extrem feingranulare Beobachtung und Erfassung über die bisher angelegten Standards hinaus erfordert,⁹ und da zum anderen diese Digitalisierung zu einem Reflexionsprozess über den analogen Vorwurf anregt, sobald bestimmte (analoge) Eigenschaften aufgegeben werden und neue (digitale) hinzukommen. Insofern ist bereits Digitalisierung und Erschließung Grundlagenforschung im besten Sinne, die in direkter Kooperation und mit Schnittstellen zwischen den Infrastruktureinrichtungen und den Forschenden erfolgreich betrieben werden kann. Bibliotheken (exemplarisch für alle kulturbe-

3 Hamburger Sonderforschungsbereich 950: Manuskriptkulturen in Asien, Afrika und Europa (<http://www.manuscript-cultures.uni-hamburg.de/>)

4 BMFB-Ausschreibungen zu „Sprache der Objekte“ (<http://www.bmbf.de/de/21609.php>); vgl. auch Foucault (1971).

5 DFG-Netzwerk „Sprachen des Sammelns“: https://www2.hu-berlin.de/sprachen_des_sammelns/Sammlungen

6 <http://www.tei-c.org/>; s. z. B. Modul 10 „Manuscript Description“; ferner die SIGs „Manuscripts; Text and Graphics“ u. a.; vgl. auch Brüning/Henzel/Pravida (2013).

7 Aktivitäten der Universität Göttingen: <http://www.uni-goettingen.de/de/419.html>; s. insbesondere auch den Katalog *Dinge des Wissens* (2012).

8 Vgl. hierzu auch den Beitrag von Elmar Mittler in diesem Band (s. S. 15 ff., insbes. S. 16 f.).

9 Vgl. hierzu auch den Beitrag von Gabriele Radecke, Mathias Göbel und Sibylle Söring in diesem Band (s. S. 85 ff., insbes. S. 101 f.).

währenden und auch kulturschaffenden Institutionen), die seit Jahrhunderten hochentwickelte Kulturtechniken des Bewahrens und Erschließens (auch des Aussortierens, Vernichtens und Vergessens) entwickelt haben, sind damit auch aufgerufen, gemeinsam mit den geisteswissenschaftlich Forschenden die entsprechenden digitalen Kulturtechniken zu entwickeln, zu pflegen und zu vermitteln.

Durch die Digitalisierung der analogen Infrastruktur ändern sich sowohl die Rahmenbedingungen als auch die Kulturtechniken der Forschung in den Geisteswissenschaften demnach dramatisch – so wie in der Physik durch „mehr Energie“ auch neue und mehr Erkenntnisse erzielt wurden und noch zu erwarten sind,¹⁰ so bedeutet der Zugang zu bislang nicht zugänglichen Quellen oder auch der Zugriff auf „mehr Quellen“ für die Geisteswissenschaften neue und mehr Erkenntnisse. Gerade wenn wir die Geisteswissenschaften als Wissenschaften vom Verstehen apostrophieren, wird deutlich, wie notwendig die Informationen über und der freie Zugang zu den Quellen, auf denen unsere Erkenntnisse beruhen – dass es und wo es sie gibt und wie andere Forschende sie in ihrem jeweiligen Wissenskonzext verstanden haben –, sind.

Mit der Quantitätssteigerung ist damit in jedem Fall auch ein Qualitätssprung verbunden; und die Bibliotheken können neben der Rolle als Infrastrukturfleger auch die wichtige Rolle des zuverlässigen und vertrauenswürdigen Brokers als Schnittstelle zwischen Infrastruktur und Forschenden einnehmen, da sie von beiden Bereichen „etwas verstehen“.

In der gewaltigen, z.T. erst begonnenen Aufgabe der vollständigen Digitalisierung des Kulturguts stehen Bibliotheken in Konkurrenz (aus der sich manchmal auch Partnerschaften ergeben) zu ökonomisch orientierten Anbietern.¹¹ So begrüßenswert die Verfügbarkeit der Digitalisate über diese Anbieter.

10 Vgl. den Beitrag von Arnulf Quadt in diesem Band (s. S. 355 ff.).

11 Ein Überblick über den tatsächlichen Anteil des digital Verfügbaren im Vergleich mit dem analog Verfügbaren ist schwer zu gewinnen. Der ‚Second Progress Report on the Digitisation and Online Accessibility of Cultural Material and on Digital Preservation in the European Union‘ sagt: “Empirical evidence such as the NUMERIC study (2009) indicates that the vast majority of cultural objects worth digitising are not yet digitised. For national libraries, only 3.5% of relevant cultural objects have been digitised.” Unter www.culturomics.org liest man: “The Google Labs N-gram Viewer (...) is designed to enable you to examine the frequency of words (banana) or phrases (‘United States of America’) in books over time. You’ll be searching through over 5.2

ter ist, muss auch diskutiert werden, ob und in welchem Maße eine Institution mit einem genau definierten Kulturauftrag die Kontrolle über Qualitätsstandards und (kommerzielle) Weiterverwendung ihrer Bestände aus der Hand geben sollte. Hier sollten die Forschenden die Bibliotheken durch die klare Formulierung von Qualitätsanforderungen stärken. In der häufig fehlenden Kontextualisierung, den lückenhaften Metadaten und der Missachtung rechtlicher Probleme „kommerzieller“ Angebote liegt nämlich auch die Chance der Bibliotheken als qualitätssichernde Institutionen und Partner der Forschung auf Augenhöhe, die mit der Wahrung des öffentlichen Auftrags befasst sind und Verfügbarmachung jenseits des Mainstreams gewährleisten können (und müssen!). Ebenso zentral ist der nachhaltig gewährleistete barrierefreie, ungehinderte Zugang zur Gesamtheit des analogen und digitalen kulturellen Erbes jenseits „künstlicher“ Länder-, Institutionen- oder Jahrhundertgrenzen hinweg – eine Anforderung, die Infrastruktureinrichtungen, Forschende und Politik insbesondere im digitalen Medium gemeinsam angehen müssen. Die mit der jetzigen – in aller Regel Partikularinteressen und ökonomischen Möglichkeiten folgenden – Digitalisierungspraxis verbundene Selektion und deren Folgen für das Kulturgut und für dessen Erforschung können kaum überblickt werden, daher ist hier dringender institutionenübergreifender Forschungsbedarf zu konstatieren.

Bei aller Vorsicht und aller Kritik und aller Heterogenität hat die Menge der digital verfügbaren geisteswissenschaftlichen Forschungsobjekte nun erstmals eine kritische Masse erreicht, die es erlaubt, darauf basierende spezifische Forschungsfragen zu entwickeln und anzugehen, qualitative und quantitative genuin digitale Methoden und Kulturtechniken zu entwickeln und zu verfeinern. Da dies nur gelingen kann durch den Zugriff auf die entsprechend aufbereiteten Quellen, benötigen die Bibliotheken zwangsläufig den Anschluss an diese Forschungen, denn die Methoden und Werkzeuge des Zugriffs haben Rückwirkungen auf das erforderliche Angebot, z.B. hinsichtlich Formaten, Qualitäten und Granularitäten der Digitalisate. Langfristigkeit und Nachhaltigkeit sind in diesem Zusammenhang so selbstverständliche Voraussetzungen, dass sie „eigentlich“ keiner weiteren Ausführungen bedürfen.

Um den Kreis zu schließen, muss angeführt werden, dass sich auch für das digitale Publizieren der Forschungsergebnisse – in direkter Vernetzung oder losem Zusammenhang mit den Quellen – und deren langfristiger Siche-

million books: ~4% of all books ever published!” (www.culturomics.org/Resources/A-users-guide-to-culturomics).

rung die Bibliothek anbietet: Sie wird zur Organisatorin des gesamten publizierten Forschungsprozesses in kollaborativen Arbeitsumgebungen.

Wie Ulrich Meyer-Dörpinghaus beim Festkolloquium darlegte,¹² halten Bibliotheken und Rechenzentren Angebote zur Ausbildung und Stärkung der Informationskompetenz bereit, die jedoch zu wenig ins Studium integriert seien (nicht ECTS-fähig). Ich stimme damit überein, dass diese zentrale Kompetenz der Wissensgesellschaft nicht auf spezialisierte Studiengänge, in denen das Digitale von Anfang an Schwerpunkt ist, beschränkt sein darf. In einem weiteren Verständnis als Methodenkompetenz der Forschung muss sie auch in die traditionellen Studiengänge disziplinscharfer Ausrichtung hineingetragen werden, damit digitales Forschen in der Mitte des jeweiligen Faches ankommt und Teil einer souveränen geisteswissenschaftlichen Methodenintegration wird. Besonders wichtig sind diese Facetten der Informations- und Methodenkompetenz auch in den Lehramtsstudiengängen als Teil einer modernen LehrerInnenbildung, sodass *digital literacy* fundamentaler Bildungsbestandteil werden kann. Bibliotheken, die technische, kulturelle, wissenschaftliche, kompetenzvermittelnde und soziale Infrastruktureinrichtungen sind, können auf diese Weise Partner von Bildung und Weiterbildung sein.

Abschließend möchte ich festhalten, dass es interessant ist zu sehen, dass gerade die „Infrastrukturprojekte“ TextGrid¹³, DARIAH¹⁴, CLARIN¹⁵ und EHRI¹⁶ für Bewegung und Fortschritt in der geisteswissenschaftlichen Forschung gesorgt haben, indem sie Brücken gebaut haben zwischen Universitäten, Wissenschaftsakademien, (kleineren und größeren) außeruniversitären Forschungsinstituten und Infrastruktureinrichtungen (Bibliotheken, Archiven, Rechenzentren). Eine fruchtbare Zusammenarbeit auf Augenhöhe, eine gemeinsame Sprache und Verständnis für die Perspektiven des anderen konnten sich entwickeln. Diese Forschungsinfrastrukturen sind einerseits forschunggetrieben, andererseits haben *sie* die Forschung angetrieben – eine Evolution zu beiderseitigem Nutzen.

12 Vgl. S. 229 ff. in diesem Band.

13 www.textgrid.de

14 www.de.dariah.eu

15 www.clarin.eu

16 www.ehri-project.eu

Literaturverzeichnis

- BMBF (Hrsg.) (2007): Jahr der Geisteswissenschaften – ABC der Menschheit. http://www.abc-der-menschheit.de/coremedia/generator/wj/de/03__Geisteswissenschaften/Geisteswissenschaften.html.
- BMBF (Hrsg.) (2013): Ausschreibung des Bundesministeriums für Bildung und Forschung zu „Sprache der Objekte“. <http://www.bmbf.de/de/21609.php>.
- Brüning, Gerrit; Henzel, Katrin; Pravida, Dietmar (2013): Multiple Encoding in Genetic Editions: The Case of ‘Faust’. In: *Journal of the Text Encoding Initiative* 4 (March 8, 2013). <http://jtei.revues.org/697>.
- Busa SJ, Father Roberto (2004): Foreword: Perspectives on the Digital Humanities. In: Susan Schreibman; Ray Siemens; John Unsworth (Hrsg.): *A Companion to Digital Humanities*. Oxford: Blackwell. Online: <http://www.digitalhumanities.org/companion/>.
- DFG (2013): DFG-Netzwerk „Sprachen des Sammelns“. https://www2.hu-berlin.de/sprachen_des_sammelns/Sammlungen.
- Dinge des Wissens* (2012): *Die Sammlungen, Museen und Gärten der Universität Göttingen*. Begleitbuch zur gleichnamigen Ausstellung in der Paulinerkirche Göttingen vom 2.6. bis 7.10.2012 anlässlich des 275. Jubiläums der Georg-August-Universität Göttingen, hrsg. v. der Georg-August-Universität Göttingen. Göttingen: Wallstein.
- Fish, Stanley (2012): The Digital Humanities And The Transcending Of Mortality. (January 9, 2012). <http://opinionator.blogs.nytimes.com/2012/01/09/the-digital-humanities-and-the-transcending-of-mortality>.
- Foucault, Michel (1971): *Die Ordnung der Dinge. Eine Archäologie der Humanwissenschaften (Les mots et les choses, 1966)*. Aus dem Französischen von Ulrich Köppen. Frankfurt am Main: Suhrkamp
- Gippert, Jost (2010): Digital Humanities. Was kommt ans Licht, wenn Texte und Bilder digital analysiert werden? Digital Humanities – die empirische Wende in den Geisteswissenschaften. In: *Forschung Frankfurt* 28 (2010): 3, 21–25. urn:nbn:de:hebis:30-87442.
- Schreibman, Susan; Siemens, Ray; Unsworth, John (2004): The Digital Humanities and Humanities Computing: An Introduction. In: dies. (Hrsg.): *A Companion to Digital Humanities*. Oxford: Blackwell. Online: <http://www.digitalhumanities.org/companion/>.
- Thaller, Manfred (2012): The Cologne Dialogue on Digital Humanities 2012. Controversies around the Digital Humanities: Proceedings of a Workshop. In: *Histo-*

rical Social Research Vol. 37, No. 3. Online: <http://www.cceh.uni-koeln.de/events/CologneDialogue>.

Wissenschaftsrat (2005): Empfehlungen zur Entwicklung und Förderung der Geisteswissenschaften in Deutschland. <http://www.wissenschaftsrat.de/download/archiv/geisteswissenschaften.pdf>.

Aus Sicht der Naturwissenschaften: Grid-Computing in der Teilchenphysik, Fortschritt durch Technik

von Arnulf Quadt

Zunächst einmal danke ich den OrganisatorInnen der Festveranstaltung für die Einladung und gratuliere der Abteilung Forschung & Entwicklung ganz herzlich zum zehnjährigen Bestehen und zu den Errungenschaften der letzten Jahre, von denen auch die Naturwissenschaften in Göttingen stark profitiert haben. Auch dafür ganz herzlichen Dank!

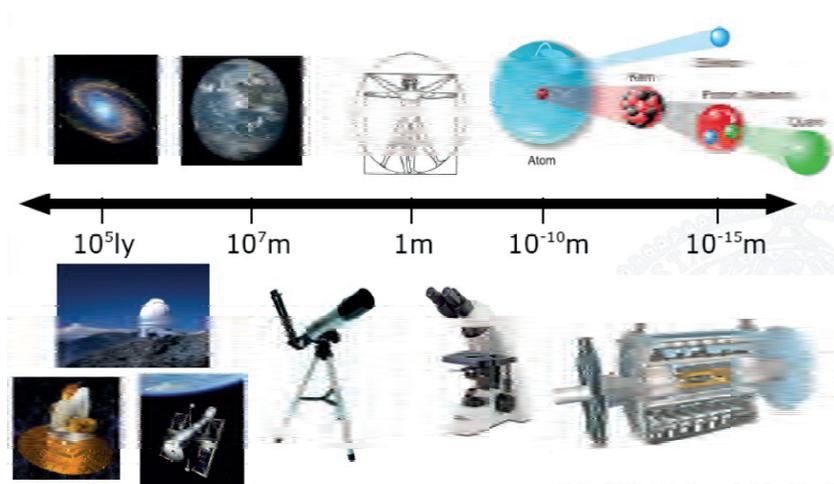


Abb. 1 Physik: Infrastrukturverbesserungen ermöglichen neue Forschungsfragen

In diesem Impulsvortrag wird am Beispiel der Teilchenphysik gezeigt, wie Verbesserungen der Methodik und insbesondere der Infrastruktur einer Forschungsrichtung neue Einblicke und Erkenntnisse ermöglichen und bedingen. Die moderne Physik widmet sich den Objekten der Natur von den ganz großen Strukturen wie dem gesamten Kosmos mit seinen Galaxien, Sonnensystemen und Planeten über Objekte in der Größe des Menschen bis hin zu den Atomen und deren elementaren Bausteinen, nämlich den Elektronen und den Atomkernen bzw. Protonen und Neutronen, die ihrerseits aus Quarks bestehen. Diese Objekte der Natur – von den größten bis zu den

kleinsten Dimensionen über zig Größenordnung verteilt – lassen sich von Erd- oder Satelliten-gestützten Teleskopen, Mikroskopen bzw. den modernen Teilchenphysik-Beschleunigern und Detektoren untersuchen. Es sind diese hochentwickelten und mit viel Erfahrung optimierten Messinstrumente, die die Einblicke in die Natur überhaupt erst ermöglichen.

Schaut man sich die elementaren Bausteine der Natur, die Leptonen (das Elektron ist ein Lepton) und Quarks genauer an, stellt man fest, dass sie jeweils in drei Generationen von Teilchenpaaren auftreten, die von der ersten über die zweite bis zur dritten Generation in ihrer Masse sehr stark anwachsen. Bis heute ist das Konzept der (trägen) Masse unverstanden. Insbesondere das beobachtete Muster großer und kleiner Massen ist heute und bis auf weiteres ein Rätsel, das die PhysikerInnen fasziniert. Während die Quarks der ersten Generation in den 60er-Jahren und die der zweiten Generation Anfang der 70er-Jahre entdeckt wurden, wurde das Bottom-Quark der dritten Generation erst Ende der 70er-Jahre, das Top-Quark als sein deutlich schwereres Partnerteilchen erst 1995 experimentell entdeckt. Eine ähnliche historische Entwicklung zeigt sich bei den Neutrinos. Ihre Existenz wurde in den 30er-Jahren als Erklärungshypothese zur Rettung der Energieerhaltung eingeführt. Das Neutrino der ersten Generation wurde in den 50er-Jahren experimentell nachgewiesen, das der zweiten Generation in den 60er-Jahren und das Tau-Neutrino der dritten Generation erst im Jahr 2000. Bei genauer Betrachtung stellt man sehr schnell fest, dass der zeitliche Ablauf dieser großen Entdeckungen und wissenschaftlichen Durchbrüche eng mit dem exponentiellen Anstieg der in Teilchenbeschleunigern erreichbaren Energie („Livingston plot“) oder mit signifikanten Neuerungen und Fortschritten in der Entwicklung präziser Teilchendetektoren zum Teilchennachweis verbunden ist. Verbesserte Forschungsinfrastruktur ermöglicht wissenschaftliche Durchbrüche und signifikant neue Erkenntnisse.

Eine ähnliche Revolution erleben wir gerade mit den Experimenten am Large Hadron Collider (LHC) des europäischen Zentrums für Teilchenphysik CERN in Genf. Mit einer aktuellen Schwerpunktsenergie von 7 TeV, die in den nächsten Jahren auf 14 TeV ausgebaut werden wird, ist dieser Proton-Proton-Collider mit Abstand der höchstenergetische und damit der leistungsstärkste Beschleuniger der Welt. 40 Millionen Mal pro Sekunde treffen mehrere Milliarden Protonen aufeinander und erzeugen Kollisionsereignisse, deren Produkte aus Hunderten oder Tausenden von Teilchen bestehen und von den vier Experimenten ATLAS, CMS, ALICE und LHC-B aufgezeichnet werden.

Diese Art der Forschung gleicht buchstäblich der Suche nach der Nadel im Heuhaufen. Während die Auswertung der aufgezeichneten Teilchenkollisionen in den 60er- und 70er-Jahren noch an Projektionstischen manuell erledigt wurde, ist heute die Datenauswertung im Vergleich zu Simulationsrechnungen nur noch bei weltweit verteilter Datenspeicherung und der Nutzung von weltweit verteilten CPU-Ressourcen möglich. Dafür wurde in den vergangenen zehn Jahren das Konzept des Grid-Computings entwickelt und bis zur Anwendungsreife verfeinert. Ähnlich wie die Entwicklung des World Wide Web in den 90er-Jahren am CERN ist auch das Grid-Computing ein neues Werkzeug, das sehr schnell außerhalb der Teilchenphysik Anwendung fand und völlig neuartige Forschungsmethoden erlaubt. Beste Beispiele dafür sind die Projekte MediGrid und TextGrid, an denen federführend Göttinger WissenschaftlerInnen, insbesondere die Abteilung Forschung & Entwicklung der SUB, beteiligt sind bzw. waren. Durch den interdisziplinären Zusammenschluss aller am Grid-Computing interessierter und beteiligter WissenschaftlerInnen und den daraus resultierenden gemeinsamen Betrieb der Speicher- und CPU-Ressourcen im GoeGrid-Cluster ist in Göttingen ein hocheffizientes Grid-Computing-Cluster entstanden, das national und international besondere Sichtbarkeit erhalten hat, in dieser Form einzigartig ist und einen signifikanten Beitrag zur Analyse der LHC-Daten aus den Jahren 2010 bis 2012 geliefert hat.

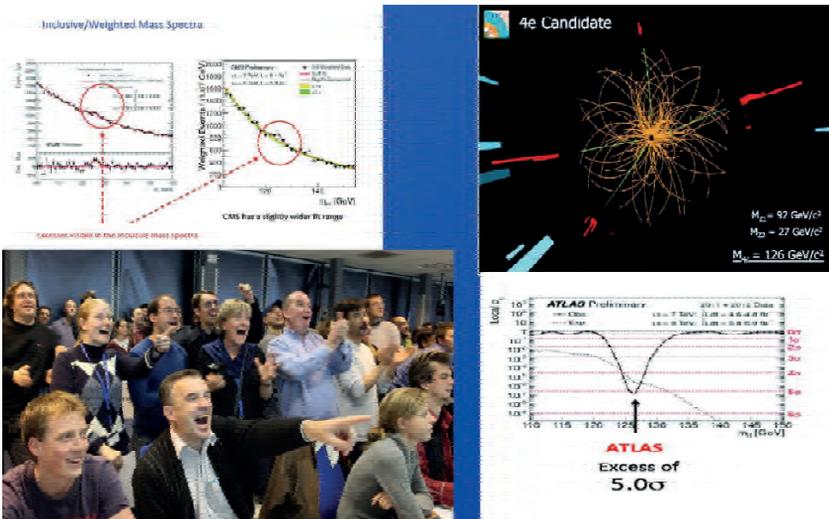


Abb. 3 Ergebnis: Die Entdeckung des Higgs-Bosons, 2012

Darauf basierend wurde am 4. Juli 2012 in einem live ausgestrahlten Seminar am CERN durch die beiden Experimente ATLAS und CMS die Entdeckung eines neuen Bosons verkündet, das sich in allen bisher untersuchten Eigenschaften als kompatibel mit dem seit 1964 gesuchten Higgs-Boson darstellt. Weitere Untersuchungen werden aktuell mit Nachdruck durchgeführt bzw. in Zukunft folgen. Auch diese großartige Entdeckung war nur durch den Einsatz neuer Detektortechnologien und verbesserte Forschungsinfrastruktur in Form des Large Hadron Colliders bzw. das neue Konzept des Grid-Computings möglich.

Erst die verbesserte Forschungsinfrastruktur führt zur Beantwortung wissenschaftlicher Fragestellung und zum Entstehen neuer Fragen. Auch das ist eine wissenschaftliche Erkenntnis, die jedoch nicht neu ist, sondern sich in regelmäßigen Abständen immer wiederholt, sowohl in den Natur- als auch in den Geisteswissenschaften. Nochmals herzlichen Glückwunsch der SUB-Abteilung Forschung & Entwicklung für die Errungenschaften und Beiträge in den letzten zehn Jahren zu eben diesen Fortschritten und alles Gute für die Zukunft!

Verzeichnis der Autorinnen und Autoren

Name	Institution
Dr. Tobias Blanke	King's College London
Dr. Harry Enke	Leibniz-Institut für Astrophysik Potsdam
Michal Frankl	Institute for War, Holocaust and Genocide Studies, Amsterdam (NIOD)
Christiane Fritze	Niedersächsische Staats- und Universitätsbibliothek Göttingen
Dr. Bernadette Fritzschn	Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar- und Meeresforschung, Bremerhaven
Stefan E. Funk	Niedersächsische Staats- und Universitätsbibliothek Göttingen
Peter Gietz	Directory Applications for Advanced Security and Information Management (DAASI International)
Mathias Göbel	Niedersächsische Staats- und Universitätsbibliothek Göttingen
Prof. Dr. Axel Horstmann	Universität Hamburg
Prof. Dr. Fotis Jannidis	Universität Würzburg
Martina Kerzel	Niedersächsische Staats- und Universitätsbibliothek Göttingen
Dr. Jens Klump	Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum
Dr. Conny Kristel	Institute for War, Holocaust and Genocide Studies, Amsterdam (NIOD)
Prof. Dr. Norbert Lossau	Georg-August-Universität Göttingen, Niedersächsische Staats- und Universitätsbibliothek Göttingen
Jens Ludwig	Niedersächsische Staats- und Universitätsbibliothek Göttingen
Dr. Ulrich Meyer-Doeringhaus	Hochschulrektorenkonferenz, Bonn
Prof. em. Dr. Dr. h.c. mult. Elmar Mittler	Georg-August-Universität Göttingen

Dr. Heike Neuroth	Niedersächsische Staats- und Universitätsbibliothek Göttingen
Prof. Dr. Achim Oßwald	Fachhochschule Köln
Wolfgang Pempe	Deutsches Forschungsnetz (DFN)
Prof. Dr. Arnulf Quadt	Georg-August-Universität Göttingen
Dr. Gabriele Radecke	Georg-August-Universität Göttingen
Prof. Dr. Andrea Rapp	Technische Universität Darmstadt
Prof. Dr. Malte Rehbein	Universität Passau
Dr. Mike Reich	Georg-August-Universität Göttingen
Kepa J. Rodriguez	Niedersächsische Staats- und Universitätsbibliothek Göttingen
Dr. Patrick Sahle	Universität zu Köln
Dr. Matthias Schulze	Universitätsbibliothek Stuttgart
Dr. Kathleen Smith	Niedersächsische Staats- und Universitätsbibliothek Göttingen
Sibylle Söring	Niedersächsische Staats- und Universitätsbibliothek Göttingen
Ph.D. Reto Speck	King's College London
Ralf Stockmann	Niedersächsische Staats- und Universitätsbibliothek Göttingen / Staatsbibliothek Berlin
Stefan Strathmann	Niedersächsische Staats- und Universitätsbibliothek Göttingen
Dr. Veerle Vanden Daelen	Universiteit Antwerpen
Ubbo Veentjer	Niedersächsische Staats- und Universitätsbibliothek Göttingen
Thorsten Vitt	Universität Würzburg
Dr. Heiko Weber	Akademie der Wissenschaften zu Göttingen
Prof. Dr. Werner Wegstein	Universität Würzburg
Dipl.-Ing. Peter Wittenburg	Max Planck Institute for Psycholinguistics, Nijmegen

Abkürzungsverzeichnis

ADHO	Alliance of Digital Humanities Organizations
AdWG	Akademie der Wissenschaften zu Göttingen
ALLEA	ALL European Academies, The European Federation of National Academies of Sciences and Humanities
ARIADNE	Advance Research Infrastructure for Archaeological Dataset Networking
ASTROCAT	ASTROnomical CATalog
BBAW	Berlin-Brandenburgische Akademie der Wissenschaften
BMBF	Bundesministerium für Bildung und Forschung
BMBFT	Bundesministerium für Forschung und Technologie
C3Grid	Collaborative Climate Community Data and Processing Grid
CDRS	Center for Digital Research and Scholarship
CeGIM	Zentrum für geowissenschaftliches Informationsmanagement
CENDARI	Collaborative European Digital Archive Infrastructure
CERL	Consortium of European Research Libraries
CERN	Conseil Européen pour la Recherche Nucléaire (Europäischer Rat für Nuklearforschung)
CHI	Collection-Holding Institutions
CLARIN	Common Language Resources and Technology Infrastructure
CMDI	Component Metadata Infrastructure
CRUD	Create, Retrieve, Update and Delete
D-GRID	German Grid Initiative
DARIAH	Digital Research Infrastructure for the Arts and Humanities
DASISH	Data Service Infrastructure for the Social Sciences and Humanities
DC	Dublin Core Metadata Element Set
DDB	Deutsche Digitale Bibliothek
DFG	Deutsche Forschungsgemeinschaft
DFN-AAI	Deutsches Forschungsnetz – Authentifizierungs- und Autorisierungs-Infrastruktur

DGI	Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis e.V.
DH	Digital Humanities
DHSI	Digital Humanities Summer Institute
DIEPER	Digitised European Periodicals
DigCurV	Digital Curator Vocational Education Europe
DINI	Deutsche Initiative für Netzwerkinformation e.V.
DKRZ	Deutsches Klimarechenzentrum
DLF	Digital Library Framework
DMV	Deutsche Mathematische Vereinigung
DNS	Domain Name System
DOI	Digital Object Identifier
DPE	Digital Preservation Europe
DTA	Deutsches Textarchiv
DTD	Document Type Definitions
DWB	Deutsches Wörterbuch von Jacob und Wilhelm Grimm
ECF	Eclipse Communication Framework
ECTS	European Credit Transfer and Accumulation System
EHRI	European Holocaust Research Infrastructure
EMANI	Electronic Mathematical Archiving Network Initiative
EPIC	European Persistent Identifier Consortium
ERIC	European Research Infrastructure Consortium
ESFRI	European Strategy Forum on Research Infrastructures
ESS	European Social Survey
EZOLO	Early Zoological Literature Online
fMRI	funktionelle Magnetresonanztomografien
GATE	General Architecture for Text Engineering
GBV	Göttinger Bibliotheksverbund
GCDH	Göttingen Centre for Digital Humanities
GDZ	Digitalisierungszentrum der Niedersächsischen Staats- und Universitätsbibliothek Göttingen
GFZ	GeoForschungsZentrum Potsdam

GI	Gesellschaft für Informatik
GND	Gemeinsame Normdatei der Deutschen Nationalbibliothek
GSW	Geistes- und Sozialwissenschaften
GT	Globus Toolkit
GWGD	Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen
GWK	Gemeinsame Wissenschaftskonferenz des Bundes und der Länder
GZG	Geowissenschaftliches Zentrum der Universität Göttingen
HEFCE	Higher Education Funding Council for England
HEP	Hochenergiephysik
HLEG	High Level Expert Group on Digital Libraries
HPC	High-Performance Computing
HRK	Hochschulrektorenkonferenz
IBI	Institut für Bibliotheks- und Informationswissenschaft Berlin
ICSU	International Council for Science
IDE	Institut für Dokumentologie und Editorik
IDF	International DOI Foundation
IETF	Internet Engineering Task Force
IGY	International Geophysical Year
INF	Infrastrukturprojekte im Rahmen von SFBs
INRIA	Institut national de recherche en informatique et en automatique (französische Forschungseinrichtung)
IPCC	Intergovernmental Panel on Climate Change
ISAD(G)	International Standard Archival Description (General)
ISDIAH	International Standard for Describing Institutions with Archival Holdings
ITASIA	Information, Technology and Society in Asia
JISC	Joint Information Systems Committee
JMP	Jewish Museum in Prague
JSTOR	Journal STORage
KII	Kommission „Zukunft der Informationsinfrastruktur“
koLibRI	kopal Library for Retrieval and Ingest

KOPAL	Kooperativer Aufbau einer Langzeitarchivs digitaler Informationen
LGPL	GNU Lesser General Public License
LHC	Large Hadron Collider (Teilchenbeschleuniger)
LIDO	Lightweight Information Describing Objects
LZA	Langzeitarchivierung
MARUM	Zentrum für Marine Umweltwissenschaften Bremen
MD	Metadaten
METS	Metadata Encoding & Transmission Standard
MODS	Metadata Object Description Schema
MoU	Memorandum of Understanding
NESTOR	Network of Expertise in long-term Storage (Kooperationsverbund zum Thema „Digitale Langzeitarchivierung“)
NIOD	Nederlands Instituut voor Oorlogsdocumentatie (Niederländisches Institut für Kriegsdokumentation)
OAI/PMH	Open Archives Initiative – Protocol for Metadata Harvesting
OAIS	Open Archival Information System
ODD	One Document Does it All
OECD	Organisation for Economical Cooperation and Development
OPAC	Online Public Access Catalog
OWL	Web Ontology Language
PERICLES	Promoting and Enhancing Reuse of Information throughout the Content Lifecycle taking account of Evolving Semantics
PID	Persistent Identifier
RAW	Rohdatenformat
RDF	Resource Description Framework
RELAX NG	Regular Language Description for XML New Generation
REST	Representational State Transfer (Programmierparadigma für Webanwendungen)
RiR	Relationen im Raum
RusDML	Russian Digital Mathematics Library
SADE	Skalierbare Architektur für Digitale Editionen
SAML	Security Assertion Markup Language

SFB	Sonderforschungsbereich
SGML	Standard Generalized Markup Language
SHAMAN	Sustaining Heritage Access through Multivalent Archiving
SHARE	Survey of Health, Ageing and Retirement in Europe
SIG TEI	TEI Special Interest Group
SOEB	Sozioökonomische Berichterstattung
SPARQL	SPARQL Protocol and RDF Query Language
SSH	Social Science and Humanities
SSO	Single Sign On
SUB	Niedersächsische Staats- und Universitätsbibliothek Göttingen
TCP	Transmission Control Protocol
TEI	Text Encoding Initiative
TERENA	Trans-European Research and Education Networking Association
TextGridRep	TextGrid-Repositories
TG	TextGrid
TIB	Technische Informationsbibliothek
TIFF	Tagged Image File Format
TMF	Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V.
UB	Universitätsbibliothek
URI	Uniform Resource Identifier
VCC	Virtual Competency Centre (virtuelles Kompetenzzentrum)
ViRR	Virtueller Raum Reichsrecht
VRE	Virtual Research Environment
VO	Virtual Observatory
W3C	World Wide Web Consortium
WDC	World Data Center
WDS	World Data System
WLCG	Worldwide Large Hadron Collider (LHC) Computing Grid
WR	Wissenschaftsrat

XML	eXtensible Markup Language
XMPP	Extensible Messaging and Presence Protocol
XSL-FO	eXtensible Stylesheet Language Formatting Objects
XSLT	eXtensible Stylesheet Language Transformation
YIVO	Yidisher Visnshaftlekher Institut (Jüdisches Wissenschaftliches Institut)
ZVDD	Zentrales Verzeichnis Digitalisierter Drucke

Weitere Titel aus dem vwh-Verlag (Auszug)

nestor Handbuch Eine kleine Enzyklopädie der digitalen Langzeitarchivierung
Version 2.0, hrsg. v. Heike Neuroth, Achim Oßwald, Regine Scheffel, Stefan Strathmann, Mathias Jehn

Juni 2009, ISBN 978-3-940317-48-3, 620 S., 24,90 €(D), 25,60 €(A), 37,90 CHF

Eine ausgewählte Pressestimme: *Rundbrief Fotografie* 16 (2009), Heft 4, S. 30–32: *Die Sammlung an Kompetenz ist beachtenswert, das Thema dringlich. Man kann sich nur wünschen, dass mit der Herausgabe dieses Handbuches auch in der Öffentlichkeit das Bewusstsein für die Bedeutung der Bewahrung des digitalen kulturellen Erbes gefördert wird. Für direkt Betroffene und Fachleute ist es eine unerlässliche Arbeitshilfe im Kampf gegen das „Dark Age“, den Verlust digitalen Kulturerbes (oder auch „nur“ der persönlichen digitalen Daten).*

Langzeitarchivierung von Forschungsdaten Eine Bestandsaufnahme

hrsg. v. Heike Neuroth, Stefan Strathmann, Achim Oßwald, Regine Scheffel, Jens Klump, Jens Ludwig im Rahmen des Kooperationsverbundes nestor

Febr. 2012, ISBN 978-3-86488-008-7, 382 S., 29,90 €(D), 30,74 €(A), 36,90 CHF

Forschungsdaten und der langfristige Zugriff auf sie sind für Wissenschaftler aller Disziplinen von großer Bedeutung: als Nachweis des Forschungsprozesses und seiner Ergebnisse, zur Nachnutzung im Rahmen inner- und interdisziplinärer Kooperationen oder im Hinblick auf vergleichende Analysen, bei denen mit neuen Methoden oder unter veränderten Rahmenbedingungen geforscht wird. Dies alles ist nur möglich, wenn Forschungsdaten gesichert, für eine langfristige Nachnutzung archiviert und zur Verfügung gestellt werden. Angesichts rasant anwachsender digitaler Datenmengen ist die Langzeitarchivierung von Forschungsdaten für alle Wissenschaftsdisziplinen eine begleitende Infrastrukturaufgabe. In dieser Untersuchung haben WissenschaftlerInnen aus elf Fachdisziplinen – Geisteswissenschaften, Sozialwissenschaften, Psycholinguistik, Pädagogik, Altertumswissenschaft, Geowissenschaft, Klimaforschung, Biodiversität, Teilchenphysik, Astronomie und Medizin – systematisch den Stand und die aktuelle Praxis im Umgang mit der Langzeitarchivierung von Forschungsdaten in ihrer jeweiligen Disziplin aufgearbeitet.

Leitfaden zum Forschungsdaten-Management

Handreichungen aus dem WissGrid-Projekt hrsg. von Jens Ludwig und Harry Enke
Jan. 2013, Hardcover, 120 S., Abb. S/W, ISBN 978-3-86488-032-2, 15,80 €

Digitale Forschungsdaten sind eine unverzichtbare Grundlage moderner Wissenschaft. Mit ihnen sind eine Reihe von Datenmanagement-Fragen verbunden: Wie lange sollen die Daten aufbewahrt werden? Welche Kontextinformationen müssen erfasst werden, um die Daten zu einem späteren Zeitpunkt noch sinnvoll benutzen zu können? Wie viel kostet die Aufbewahrung? Das Buch stellt mit einem Leitfaden und einer Checkliste einfach handhabbare Instrumente für fachwissenschaftliche oder Infrastruktur-Einrichtungen bereit, um die wichtigsten Aufgaben und Fragen im Forschungsdaten-Management strukturiert beantworten und effizient planen zu können.

Digital Curation of Research Data Experiences of a Baseline Study in Germany

hrsg. von Heike Neuroth, Stefan Strathmann, Achim Oßwald, Jens Ludwig
Nov. 2013, Broschur, 94 S., ISBN 978-3-86488-054-4, 12,80 €

Bei dieser englischsprachigen Broschüre handelt es sich um eine Kurzfassung der umfangreichen deutschsprachigen Publikation *Langzeitarchivierung von Forschungsdaten* (s.o.) vom Februar 2012.

Reihe „Schriften zur Informationswissenschaft“

M. Heckner: Tagging, Rating, Posting
2009, 27,90 € ISBN 978-3-940317-39-1

S. Mühlbacher:

Information Literacy in Enterprises
2009, 32,90 € ISBN 978-3-940317-45-2

M. Maßun: Collaborative Information
Management in Enterprises
2009, 28,90 € ISBN 978-3-940317-49-0

T. Memmel: User Interface Specification
for Interactive Software Systems
2009, 33,90 € ISBN 978-3-940317-53-7

A. Ratzka: Patternbasiertes User Inter-
face Design für multimodale Interak-
tion 2010, 33,90 € ISBN 978-3-940317-62-9

M. Prestipino: Die virtuelle Gemein-
schaft als Informationssystem
2010, 30,90 € ISBN 978-3-940317-69-8

A. Warta: Kollaboratives Wissens-
management in Unternehmen
2011, 30,90 € ISBN 978-3-940317-90-2

J. Griesbaum, T. Mandl, C. Womser-
Hacker (Hg.): Information und Wis-
sen: global, sozial und frei?
Proceedings ISI 2011

2011, 36,50 € ISBN 978-3-940317-91-9

M. Görtz: Social Software as a Source
of Information in the Workplace
2011, 31,90 € ISBN 978-3-86488-006-3

C. Carstens: Ontology Based Query
Expansion Retrieval Support for the
Domain of Educational Research
2012, 34,90 € ISBN 978-3-86488-011-7

A. Köstlbacher: Eine Informations-
system-Infrastruktur für das Thera-
peutische Drug Monitoring
2012, 27,90 € ISBN 978-3-86488-019-3

S.-J. Untiet-Kepp: Adaptive Feedback
zur Unterstützung in kollaborativen
Lernumgebungen
2012, 30,90 € ISBN 978-3-86488-023-0

H.-C. Hobohm (Hg.): Informations-
wissenschaft zwischen virtueller Infra-
struktur und materiellen Lebenswelten
Proceedings des 13. Internationalen Sym-
posiums für Informationswissenschaft (ISI
2013), Potsdam, 19.–22. März 2013
2013, 34,50 € ISBN 978-3-86488-035-3

Reihe „E-Humanities“

G. Vogl: Selbstständige Medienschaf-
fende in der Netzwerkgesellschaft
2008, 29,90 € ISBN 978-3-940317-38-4

C. Russ: Online Crowds Massenphäno-
mene und kollektives Verhalten im Inter-
net 2010, 31,50 € ISBN 978-3-940317-67-4

C. Potzner: Chancen und Risiken
der Arbeit im E-Business
2010, 29,90 € ISBN 978-3-940317-70-4

M. Janneck/C. Adelberger: Komplexe
Software-Einführungsprozesse gestal-
ten: Grundlagen und Methoden Am
Beispiel eines Campus-Management-Sys-
tems 2012, 26,90 € ISBN 978-3-940317-63-6

H. Kohle: Digitale Bildwissenschaft
2013, 16,80 € ISBN 978-3-86488-036-0

Reihe „Web 2.0“

J. Moskaliuk (Hg.): Konstruktion und
Kommunikation von Wissen mit Wikis
2008, 27,50 € ISBN 978-3-940317-29-2

H. Frohner: Social Tagging
2010, 26,90 € ISBN 978-3-940317-03-2

R. Bauer: Die digitale Bibliothek
von Babel Über den Umgang
mit Wissensressourcen im Web 2.0
2010, 26,90 € ISBN 978-3-940317-71-1

J. Jochem: Performance 2.0
Zur Mediengeschichte der Flashmobs
2011, 24,90 € ISBN 978-3-940317-98-8

G. Franz: Die vielen Wikipedias
Vielsprachigkeit als Zugang zu einer globali-
sierten Online-Welt
2011, 27,50 € ISBN 978-3-86488-002-5

R. Sonnberger: Facebook im Kontext
medialer Umbrüche
2012, 29,50 € ISBN 978-3-86488-009-4

J. Brailovskaia:
Narzisstisch und sensationssuchend?
Eine Studie zum Zusammenhang zwischen
Persönlichkeitsmerkmalen und Online-
Selbstdarstellung am Beispiel von *studivZ*
2013, 24,50 € ISBN 978-3-86488-039-1

Reihe „Medientheorie“

W. Drucker: Von Sputnik zu Google
Earth Über den Perspektivenwechsel hin
zu einer ökologischen Weltansicht
2011, 25,90 € ISBN 978-3-940317-82-7

S. Brugner: Über die Realität im Zeit-
alter digitaler Fotografie
2012, 23,90 € ISBN 978-3-86488-018-6

Reihe „E-Learning“

- T. Strasser: Moodle im Fremdsprachenunterricht
2011, 28,50 € ISBN 978-3-940317-92-6
- M. Nagl: Game-Assisted E-Learning in der Umweltbildung
2011, 28,50 € ISBN 978-3-940317-94-0
- D.Schirmer et al.: Studieren als Konsum
Veralltäglichung und Degendering von E-Learning
2011, 27,90 € ISBN 978-3-940317-83-4
- C. Biel: Personal Learning Environments als Methode zur Förderung des selbstorganisierten Lernens
2011, 24,90 € ISBN 978-3-86488-001-8
- A. Blessing: Personalisiertes E-Learning
2012, 29,90 € ISBN 978-3-86488-007-0
- I. Zorn: Konstruktionstätigkeit mit Digitalen Medien Eine qualitative Studie als Beitrag zur Medienbildung
2012, 36,50 € ISBN 978-3-86488-012-4
- K. Himpsl-Gutermann: E-Portfolios in der universitären Weiterbildung
2012, 30,90 € ISBN 978-3-86488-014-8
- M. Beißwenger/N. Anskait/A. Storrer (Hg.): Wikis in Schule und Hochschule
2012, 36,90 € ISBN 978-3-86488-017-9
- P. Bettinger: Medienbildungsprozesse Erwachsener im Umgang mit sozialen Online-Netzwerken
2012, 25,90 € ISBN 978-3-86488-020-9
- C. Lehr: Web 2.0 in der universitären Lehre Ein Handlungsrahmen für die Gestaltung technologiegestützter Lernszenarien
2012, 27,90 € ISBN 978-3-86488-024-7
- J. Wagner/V. Heckmann (Hg.): Web 2.0 im Fremdsprachenunterricht Ein Praxisbuch für Lehrende in Schule und Hochschule
2012, 27,50 € ISBN 978-3-86488-022-3
- E. Blaschitz et al. (Hg.): Zukunft des Lernens Wie digitale Medien Schule, Aus- und Weiterbildung verändern
2012, 23,50 € ISBN 978-3-86488-028-5
- U. Höbarth: Konstruktivistisches Lernen mit Moodle Praktische Einsatzmöglichkeiten in Bildungsinstitutionen - 3. Aufl. -
2013, 31,50 € ISBN 978-3-86488-033-9

- A. Klampfer: E-Portfolios als Instrument zur Professionalisierung in der Lehrer- und Lehrerinnenausbildung
2013, 27,90 € ISBN 978-3-86488-034-6
- M. Hielscher: Autorentools für multimediale und interaktive Lernbausteine Architektur und Einsatzszenarien von LearningApps.org
2013, 26,50 € ISBN 978-3-86488-041-4
- C. Koenig: Bildung im Netz Analyse und bildungstheoretische Interpretation der neuen kollaborativen Praktiken in offenen Online-Communities
2013, 31,90 € ISBN 978-3-86488-042-1

Reihe „Medienwirtschaft“

- K. Huemer: Die Zukunft des Buchmarktes Verlage und Buchhandlungen im digitalen Zeitalter
2010, 24,90 € ISBN 978-3-940317-73-5
- J.-F. Schrape: Gutenberg-Galaxis Reloaded? Der Wandel des deutschen Buchhandels durch Internet, E-Books und Mobile Devices
2011, 17,90 € ISBN 978-3-940317-85-8
- B. Blaha: Von Riesen und Zwergen Zum Strukturwandel im verbreitenden Buchhandel in Deutschland und Österreich
2011, 24,90 € ISBN 978-3-940317-93-3
- J. Stiglhuber: Macht und Ohnmacht der Unabhängigen Independent-Verlage und ihre Verortung
2011, 26,90 € ISBN 978-3-86488-003-2

Kleine Schriften

- J.-F. Schrape: Wiederkehrende Erwartungen Visionen, Prognosen und Mythen um neue Medien seit 1970 [Softcover]
2012, 11,90 € ISBN 978-3-86488-021-6

*weitere Reihen im vwh-Programm
(s. www.vwh-verlag.de):*

- **TypoDruck**
- **AV-Medien**
- **Game Studies**
- **Multimedia**
- **E-Business**
- **E-Collaboration**
- **Schriften des Innovators Club**



Aktuelle Ankündigungen, Inhaltsverzeichnisse und Rezensionen finden sie im vwh-Blog unter www.vwh-verlag.de.

Das komplette Verlagsprogramm mit Buchbeschreibungen sowie eine direkte Bestellmöglichkeit im vwh-Shop finden Sie unter www.vwh-verlag-shop.de.



Über dieses Buch

Die Festschrift zum zehnjährigen Bestehen der Abteilung Forschung und Entwicklung (F&E) der Niedersächsischen Staats- und Universitätsbibliothek Göttingen in 2012 resümiert entlang ausgewählter Meilensteine die Arbeit im vergangenen Jahrzehnt. Dabei begibt sie sich auf die Spuren des Wandels im Leitbild wissenschaftlicher Bibliotheken.

Gemeinsam mit Partnern aus der Wissenschaft wurden zahlreiche, wegweisende Projekte auf dem Gebiet der digitalen Forschung vorangebracht. Dieser Pionierarbeit in der deutschen Bibliothekslandschaft wird in 19 Aufsätzen von über 30 Autorinnen und Autoren nachgegangen.



ZUKUNFT



vwh Verlag Werner Hülsbusch
 Fachverlag für Medientechnik und -wirtschaft
www.vwh-verlag.de

In Kooperation mit dem



Universitätsverlag Göttingen

24,90 € (D)

25,60 € (A)

32,90 CHF

ISBN: 978-3-86488-043-8



9 783864 880438