



**Erster Bericht über die Verwendung
von Verfahren der *Digital Humanities*
in den Geistes- und Kulturwissenschaften
(R 2.2.3)**

Version 1.0 (31.08.12)

Arbeitspaket 2.2

Verantwortlicher Partner TUD

DARIAH-DE

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

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

GEFÖRDERT VOM



**Bundesministerium
für Bildung
und Forschung**

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

BMBF Förderkennzeichen: 01UG1110A bis M

Laufzeit: März 2011 bis Februar 2014

Dokumentstatus: <Final, V 1.0 >

Verfügbarkeit: <öffentlich, DARIAH-DE-intern>

Autoren:

Rainer Becker, TUD (RB)

Michael Bender, TUD (MB)

Matthew Munson, GCDH (MM)

Ruth Reiche, TUD (RR)

Stefan Schmunk, IEG (StS)

Christof Schöch, UWÜ (CS)

Inhaltsverzeichnis

1. Einleitung: Ziel des Berichts	4
2. DH-Verfahren im Forschungsprozess	6
2.1. Definition	6
2.2. DH-Aktivitäten der Digitalisierung und Archivierung sowie der Wissenschaftskommunikation	7
3. Verwendung von DH-Verfahren: Stand in den Fachdisziplinen	13
3.1 Philologien.....	13
3.2 Geschichtswissenschaft	19
3.3 Kunstgeschichte	22
3.4 Archäologie	25
3.5 Musikwissenschaften.....	27
3.6 Theologie.....	28
3.7 Philosophie.....	30
3.8 Jüdische Studien	32
4. Fazit: Kern-DH-Verfahren als disziplinenübergreifende Perspektive für die Entwicklung der DARIAH-DE-Infrastruktur	36

1. Einleitung: Ziel des Berichts

In einer zunehmend digital durchdrungenen Welt etabliert sich der Einsatz elektronischer Medien und Technologien auch in den Geistes- und Kulturwissenschaften. Dennoch besteht weiterhin Aufholbedarf im Vergleich zu anderen Wissenschaftsgebieten. Schließlich haben die Geistes- und Kulturwissenschaften neue Materialien und Perspektiven zu bieten sowie ein anderes, neues Reflexions-"niveau" bzw. -verhalten. Insofern können und müssen sie die digital durchdrungene Welt um ihre spezifische Perspektive bereichern.

Die Digital Humanities entwickeln sich, Quantität und Qualität digitaler Ressourcen steigen kontinuierlich und verändern bestimmte Rahmenbedingungen und Forschungspraktiken in den Geistes- und Kulturwissenschaften. Bestehende Fragestellungen können mit neuen Methoden und Verfahren untersucht, neue Fragen gestellt und bearbeitet werden. Neue digitale Arbeits-, Forschungs- und Wissensformen entstehen.

Verfahren der Digital Humanities (DH) gehen zunehmend über einen Transfer analoger in digitale Verfahren hinaus und eröffnen neue Forschungsperspektiven. Die Digital Humanities umfassen grundsätzlich alle geistes- und kulturwissenschaftlichen Disziplinen, stehen zugleich der Informatik und der Informationswissenschaft nahe und bilden einen Disziplinen übergreifenden Arbeitsbereich. Für diesen Arbeitsbereich entsteht eine gemeinsame digitale Forschungsinfrastruktur wie DARIAH-DE – sie bildet den Hintergrund dieses Reports. Dieser Report hat zum Ziel, einen ersten Überblick über derzeit in einzelnen Disziplinen verwendete *DH-Verfahren* zu erstellen und einige sich daraus ergebende Perspektiven und Möglichkeiten aufzuzeigen.

Voraussetzung für dieses Vorgehen ist die anfängliche, pragmatische Bestimmung des Begriffs *DH-Verfahren* und seine Einordnung in umfassendere DH-Forschungsprozesse. Er wird von vorgelagerten Teilprozessen, wie z.B. Digitalisierungsverfahren, sowie von nachgelagerten oder begleitenden Teilprozessen, wie z.B. Archivierung, Publikation, Kommunikation bzw. Kollaboration, abgegrenzt. Der zugrunde gelegte allgemeine Verfahrensbegriff baut zugleich auf einer Unterscheidung von einem bestimmten Methodenbegriff auf.

Im Anschluss an diese Definitionen wird skizzenhaft dargestellt, wie aktuelle DH-Verfahren bislang in einigen Disziplinen der Geistes- und Kulturwissenschaften eingesetzt werden. Der vorliegende Report ist auf diejenigen geistes- und kulturwissenschaftlichen Fächer fokussiert, in deren Zusammenhängen DARIAH-DE Informationen durch Kooperationen oder Partnerschaften zugänglich wurden bzw. der Kontakt zur und/oder die eigene Aktivität in der jeweiligen Disziplin

das Erschließen von DH-Verfahren ermöglicht hat. Hierbei wird kein Anspruch auf einen umfassenden, vollständigen Überblick erhoben. Vielmehr wird ein erster Einblick mit dem Ziel einer vorläufigen Identifikation der gegenwärtigen Bandbreite digital gestützter Forschung in den Geistes- und Kulturwissenschaften und ihren Schwerpunkten verfolgt. Der vorliegende Report versteht sich damit zugleich als Ausgangspunkt eines Prozesses, der von DARIAH-DE auch künftig weiter begleitet werden wird und der grundsätzlich darauf ausgerichtet ist, zukünftige Entwicklungen zu ergänzen und weitere geisteswissenschaftliche Disziplinen einzubeziehen. Einzelne Forschungsprojekte können zukünftig auf diese Weise ihre Erfahrungen im Umgang mit Verfahren der Digital Humanities einbringen, so dass sich die Ergebnisse dieses ersten Berichts stetig ausdifferenzieren lassen und letztlich in einer „Typologie der Forschungsprozesse“ münden werden. Diese „Typologie der Forschungsprozesse“, deren erste inhaltliche Grundlage dieser Report bildet, wird in Zukunft u.a. auf dem DARIAH-DE-Portal fortgeführt und gepflegt werden.

Auf dem skizzierten Überblick zu *DH-Verfahren* aufbauend werden zentrale *DH-Verfahren*, so genannte *Hot Spots*, in Abschnitt 3 über die verschiedenen behandelten geisteswissenschaftlichen Disziplinen hinweg identifiziert, in Abschnitt 4 untereinander abgeglichen und kurz auf ihr jeweiliges inter- bzw. transdisziplinäres Potential hin untersucht. Findet sich ein disziplinär zentrales *DH-Verfahren* in mehr als einem Fach, wird im vorliegenden Report von einem *Kern-DH-Verfahren* gesprochen. Daraus lassen sich Konsequenzen für die Entwicklung bestimmter Elemente der DARIAH-DE-Infrastruktur ziehen (technische Anforderungen, Anschlussoptionen etc.; vgl. Ende Abschnitt 4). Das im Report gewählte Vorgehen kann zugleich Hinweise darauf geben, dass solche Kern-DH-Verfahren ein breiteres Anwendungsspektrum bieten, z.B. in verschiedenen Disziplinen, also einen generischen Einsatz ermöglichen könnten. Ob allerdings eine Übertragung solcher *Kern-DH-Verfahren* in Kontexte der sie aktuell nicht oder weniger stark nutzenden Disziplinen fruchtbar sein könnte – ob also jeweils von konkreteren Desideraten o.ä. in der Anwendung dieser digital gestützten Verfahren gesprochen werden kann – muss sowohl in fachspezifischen als auch in interdisziplinären wissenschaftlichen Diskursen reflektiert werden. Ein zentrales Ziel dieses Reports ist daher auch, solche wissenschaftlichen Diskurse – in denen sowohl disziplinäre als auch interdisziplinäre Perspektiven, Möglichkeiten und Grenzen der Digital Humanities sowie mögliche Anwendungsszenarien von *DH-Verfahren* thematisiert werden – zu initiieren und erste Impulse bzw. Anhalts- und Anknüpfungspunkte zu bieten.

2. DH-Verfahren im Forschungsprozess

Um pragmatisch von *DH-Verfahren* sprechen und einen Überblick über sie geben zu können, wurde eine erste Arbeitsdefinition entwickelt, die zunächst (vgl. Abschnitt 2.1) einen Begriff wissenschaftlicher Verfahren bestimmt und ihn von einem Begriff wissenschaftlicher Methoden abgrenzt. *DH-Verfahren* werden anschließend in Analogie zu diesem Verfahrens-Begriff definiert.¹

Darüber hinaus wird auf prozedurale Voraussetzungen und Bedingungen unseres *DH-Verfahrensbegriffs* – z.B. eher vorhergehende Prozesse wie die Digitalisierung jeweiliger Forschungsgegenstände – sowie auf nachfolgende oder zwischengeschaltete Prozesse wie Archivierung und Publikationspraktiken sowie Kommunikations- und Kollaborations-Praktiken eingegangen. Solche Aktivitäten werden im vorliegenden Report von computergestützten *DH-Verfahren* der Forschung im engeren Sinne pragmatisch abgesetzt (vgl. Abschnitt 2.2).

2.1. Definition

Unter einer *Methode* kann ein systematisches und regelgeleitetes Vorgehen verstanden werden, das reflexive und praktische Elemente verbindet. Die Anwendung einer solchen Methode sorgt dafür, dass jeweils gewonnene Forschungsergebnisse als valide und verlässlich gelten können, wobei Geltungsansprüche von Methoden wie Ergebnissen immer auch Ergebnis wissenschaftlicher Aushandlungsprozesse sind. Die Charakteristika einer wissenschaftlichen Forschungsmethode werden dabei durch die Beschaffenheit derjenigen Gegenstände definiert, auf die sie jeweils gerichtet ist. Forschungsmethoden sind daher zum einen stets gegenstandszentriert, zum anderen implizieren sie einen bestimmten interpretativen Horizont.

Von der Forschungsmethode wird im vorliegenden Report pragmatisch das wissenschaftliche *Verfahren* unterschieden. Ein Verfahren ist in diesem Sinne ein planvoller, systematischer und praktischer Umgang mit Forschungsgegenständen, dem die jeweilige Forschungsmethode stets übergeordnet bleibt. Ein wissenschaftliches Verfahren, wie z.B. der Vergleich mehrerer Textfassungen oder die Auswertung von Metadaten, kommt somit immer erst vor dem Hintergrund des sinnstiftenden Horizonts der jeweils gewählten Methode zum Einsatz. Zugleich erlauben solche Methoden der Forschung die Anwendung mehrerer Verfahren – durchaus auch eine Abfolge bestimmter Verfahren – um ein sorgfältig umgrenztes Erkenntnisinteresse und damit mindestens ein verbundenes Forschungsziel zu erreichen.

¹ Wir danken Oliver Immel (TU Darmstadt), der vor Fertigstellung dieses Reports das Projekt verlassen hat, dafür, dass er für die folgenden begrifflichen Abgrenzungen einige grundlegende Vorarbeiten geleistet hat.

In Analogie hierzu verstehen wir unter einem *DH-Verfahren* im engeren Sinn einen geplanten, systematischen und praktisch wirksamen wissenschaftlichen Umgang mit *digital vorliegenden* Forschungsdaten. Eine jeweils spezifische DH-Forschungsmethode kann dementsprechend auch durch Addition mehrerer *DH-Verfahren* umgesetzt oder ermöglicht werden – umgekehrt können zugleich bestimmte *DH-Verfahren* auch durchaus im Kontext mehrerer *DH-Methoden* eingesetzt werden. Verfahren der Digital Humanities sind hierbei dann jeweils bestimmt durch digital transformierte oder bereits digital vorliegende Forschungsdaten (bzw. digitale Repräsentationen bestimmter Forschungsgegenstände), die mit Hilfe von Software weiter bearbeitet oder untersucht werden. Entscheidendes Merkmal von *DH-Verfahren* ist es dabei, die Untersuchung und den durch Erkenntnisinteressen bedingten Umgang mit solchen Forschungsdaten und Gegenstandsbezügen zu leisten (*DH-Verfahren* im engeren Sinn), wobei dies durch digitale Verfahren teils lediglich erleichtert, teils überhaupt erst ermöglicht wird. *DH-Verfahren* im engeren Sinn können hierbei grob in solche der Erfassung, Aufbereitung und Auswertung von Daten unterteilt werden, wobei eine Zuordnung bestimmter Verfahren zu mehreren Kategorien möglich ist.

2.2. DH-Aktivitäten der Digitalisierung und Archivierung sowie der Wissenschaftskommunikation

Von *DH-Verfahren* der Forschung im engeren Sinn werden Verfahren abgegrenzt, die Forschungsgegenstände digital *verfügbar machen* und/oder *archivieren*, aber keine gezielte Datenerfassung bewirken, sowie Verfahren, die nicht die Aufbereitung und Auswertung von Forschungsdaten vor dem Hintergrund bestimmter DH-Forschungsfragen und Erkenntnisinteressen und -zielen leisten, sondern die Speicherung und/oder Weiterverwendung von Forschungsdaten im Sinne von Publikations-, Kommunikations- und Kollaborationsprozessen unterstützen.

Reine Digitalisierungstechniken, wie z.B. die Erstellung von Faksimiles bzw. Retrodigitalisaten, werden daher eher dem Vorfeld von *DH-Verfahren* zugeordnet, nicht den *DH-Verfahren* der Forschung im engeren Sinn. Aus diesem Grund werden sie in Abschnitt 3 nicht im Rahmen jeder einzelnen Fachdisziplin ausführlicher erläutert, sondern hier bereits vorab in Kurzdurchgängen durch Praktiken in verschiedenen Fächern exemplarisch beschrieben. Dies gilt auch für Archivierungs-, Publikations- sowie Kommunikations- bzw. Kollaborations-Verfahren, die folglich ebenfalls noch vor dem Gesamtdurchgang durch relevante *DH-Verfahren* der im Report berücksichtigten Fachdisziplinen in Abschnitt 3 behandelt werden.

Der Schwerpunkt bei der **Digitalisierung**² und **Archivierung** von Primärtexten liegt z.B. in den *Philologien* gegenwärtig darauf, digitale Faksimiles zu erstellen und präzise Metadaten zu vergeben. So werden das Suchen und Auffinden der Materialien in einem Repositorium bzw. Katalog sichergestellt sowie die Grundlagen für digitale Editionen geschaffen. Es existiert bereits eine Anzahl von Textarchiven, die qualitativ hochwertig aufbereitete Texte und Textsammlungen bereitstellen. Hier wäre im Bereich der englischsprachigen Philologie zum Beispiel das *Oxford Text Archive* (OTA)³ zu nennen oder das *E-Text Lab*, heute online auf der Webpräsenz der University of Virginia Library unter dem Namen *Scholars' Lab*⁴ verfügbar. Solche Ressourcen dienen als Ausgangspunkt für weiterführende linguistische, literaturwissenschaftliche und kulturwissenschaftliche Forschungs- und Lehraktivitäten in den Digital Humanities und werden als Quelle für den Aufbau spezifischer Textsammlungen zu einzelnen Autoren, Epochen und Genres herangezogen.

In der *Geschichtswissenschaft* ist die Digitalisierung von Literatur, Quellen, Forschungsdaten und weiterem für die Forschung interessantem historischem Quellenmaterial ebenfalls von großer Bedeutung. Dies ist ein Aspekt, der überwiegend von Bibliotheken und Archiven im Rahmen Cultural Heritage“-bezogener Aktivitäten – also der Bewahrung und Erhaltung des kulturellen Erbes durch Digitalisierung und Retrodigitalisierung – wahrgenommen wird. Hier zeigte sich, dass die 2009 von der DFG im Rahmen des Förderprogramms *Wissenschaftliche Literaturversorgungs- und Informationssysteme* (LIS) publizierte *DFG-Praxisregel Digitalisierung*⁵ wegweisend für standardisierende Digitalisierungsverfahren war und maßgeblich dazu beigetragen hat, dass Forschungsliteratur und Quellen/Forschungsdaten nun so digital aufbereitet vorliegen, dass sie unabhängig von spezifischen Forschungsfragen nachgenutzt werden können.⁶ Ein weiterer Bereich ist auch hier die Digitalisierung von Quellen und gedruckten bzw. publizierten zeitgenössischen Werken mit dem Ziel, historisch-kritische Editionen zu erstellen.

In der *Kunstgeschichte* ist von großer Relevanz, dass Manuskripte, Zeichnungen etc. möglichst hochauflösend und farbecht digitalisiert werden. Neben der Erfassung spezifischer Daten des betreffenden Werks wie Maße, Technik etc. werden in der Regel zusätzlich Daten zu Personen,

² Als einheitliche Vorgaben im Bereich der Digitalisierung von Forschungsdaten sind fächerübergreifend die *Praxisregeln Digitalisierung* der Deutschen Forschungsgemeinschaft anzusehen.

³ <http://ota.ahds.ac.uk/>

⁴ <http://www2.lib.virginia.edu/scholarslab/>

⁵ http://www.dfg.de/download/pdf/foerderung/programme/lis/praxisregeln_digitalisierung.pdf

⁶ Aus Forscherperspektive ist zu wünschen, dass solche Praxishandbücher und -empfehlungen auch in Zukunft Grundlage der Forschungsförderung werden bzw. bleiben, so dass auf diesem Weg ein Verfahren der Standardisierung auf höchster technischer und informationstechnologischer Qualität etabliert werden kann.

Institutionen, Archivalien, Literatur und Projekten aus der Forschungsliteratur erschlossen, um zu einem möglichst dichten Datenbestand zu gelangen, der die Grundlage für die Generierung und Lösung verschiedenartiger Fragestellungen bildet. Hierbei ist als Informationssystem und Plattform für die Geisteswissenschaften ZUCCARO zu nennen, das ein Datenmodell besitzt, welches mit dem CIDOC-CRM, dem ISO-Standard für Daten zum historischen Kulturgut, kompatibel ist.⁷ Forschungsdatenbanken, die auf ZUCCARO aufbauen, sind z.B. *Lineamenta* (Forschungsdatenbank für Architekturzeichnungen)⁸ und *Ars Roma* (Forschungsdatenbank zur Malerei in Rom 1580-1630)⁹. Beliebte Bilddatenbanken bzw. -archive im deutschen Raum sind *prometheus – Das verteilte digitale Bildarchiv für Forschung und Lehre*¹⁰ sowie das *Bildarchiv Foto Marburg*¹¹, die ihren Schwerpunkt auf die Kunst und Architektur in Deutschland und Europa setzen. Darüber hinaus existieren zahlreiche Museumskataloge sowie thematische Archive wie etwa das *Cranach Digital Archive*¹², das kunsthistorische, technische sowie konservatorische Informationen zu den Gemälden Lucas Cranachs und seiner Werkstatt liefert. In der *Philosophie* dient das Herstellen von Digitalisaten von Primärtexten gegenwärtig hauptsächlich jeweils individuellen Forschungszwecken. Eher selten in gebündelter Form vorliegend, münden sie aktuell z.B. in die Eröffnung spezifischer Quellen-Datenbanken wie *Early Modern Thought Online* (EMTO)¹³ der Fernuniversität Hagen – zur Philosophie der frühen Neuzeit – sowie eine partielle Integration dieser Datenbanken untereinander, insbesondere durch die virtuelle Fachbibliothek *Sophikon* der Universität Erlangen-Nürnberg, deren umfassende Suchfunktionalität hervorzuheben ist.¹⁴ Philosophische Forschungs- und Sekundärliteratur liegt in digitaler Form gegenwärtig in recht überschaubarer Form vor. Zu nennen sind Periodika, die zwar als digitaler Volltext vorliegen, aber meist gebührenpflichtig sind, wie im Falle der Deutschen Zeitschrift für Philosophie¹⁵, oder in Einzelfällen als Open Access-eJournal vorliegen, wie z.B. das *Forum Interdisziplinäre Begriffsgeschichte* des ZfL Berlin.¹⁶ Die Lage ändert sich langsam, durch digitale Paper-Sammelstellen, z.B. die Lösung

⁷ <http://zuccaro.biblhertz.it/>

⁸ <http://lineamenta.biblhertz.it/>

⁹ <http://www.biblhertz.it/forschung/forschungsprojekte-des-instituts/arsroma/>

¹⁰ <http://prometheus-bildarchiv.de/>

¹¹ <http://www.fotomarburg.de/>

¹² <http://www.lucascranach.org/>

¹³ <http://www.sophikon.de/1223.html>

¹⁴ <http://www.sophikon.de>

¹⁵ <http://www.oldenbourg-link.com/toc/dzph/>

¹⁶ <http://www.zfl-berlin.org/forum-begriffsgeschichte-detail/items/forum-interdisziplinaere-begriffsgeschichte.html>; Außerdem: <http://www.sicetnon.org/index.php?module=PostWrap&page=http://deleuze.tausendplateaus.de/&height=1000&width=820> sowie <http://www.flusserstudies.net/>

Sammelpunkt der Universität Wien¹⁷, sowie im Preprint-Bereich¹⁸ oder durch Rezensionsportale, insbesondere durch *Kritikon* der Universität Leipzig.¹⁹

Wissenskommunikation bzw. die **Dissemination von Forschungsergebnissen und -projekten** erfolgt über alle berücksichtigte Fächer hinweg mittels digitalen Formen wie elektronischen Zeitschriften, Rezensionsplattformen, wissenschaftlichen Blogs und Informationsportalen. Hypertextuelle, also auf nicht-linear organisiertem Text basierende Plattformen (z.B. Wikis, Blogs, vernetzte Textsammlungen) ermöglichen Wissens- und Ideenaustausch sowie neue Publikationsformen. Dieser Bereich ist zugleich bedeutsam für die geisteswissenschaftliche Forschungspraxis, insbesondere unter dem Aspekt kollaborativer Forschungs- und Textproduktionspraktiken. In den Philologien gibt es bspw. in der Mediävistik und der Romanistik ausdifferenzierte und umfangreiche Portale für Fachinformationen²⁰, in der Geschichtswissenschaft mit Bedeutung weit darüber hinaus ist H-Soz-Kult zu nennen²¹. Das Zeitschriftenwesen ist immer noch von Print- oder Hybrid-Publikationen geprägt, d.h. elektronische Zeitschriften sind über Lizenzen einsehbar. Die Anzahl der elektronischen Fachzeitschriften im Open Access ist überschaubar, insgesamt ist der OpenAccess-Gedanke in den Geisteswissenschaften zwar Thema, jedoch noch wenig in der Forschungspraxis verankert. Für den Bereich der Geisteswissenschaften werden im *Directory of Open Access Journals* (DOAJ) immerhin mehrere Dutzend Zeitschriften aus dem deutschsprachigen Raum aufgeführt. Zu nennen sind hier u.a. *PhiN - Philologie im Netz*²², *IMAGE - Zeitschrift für interdisziplinäre Bildwissenschaft*²³, das *Historische Forum*²⁴ oder das *Marburg Journal of Religion*.²⁵

Im Rezensionswesen wird von Literaturwissenschaftlern im Wesentlichen *lasl-online* genutzt, zudem werden Plattformen aus anderen Disziplinen mitgenutzt, so zum Beispiel *recensio.net*, eine Plattform, die Rezensionen aggregiert. Die Praxis des wissenschaftlichen Bloggens ist in den Literaturwissenschaften wenig etabliert, auch das kürzlich gestartete geisteswissenschaftliche Blogportal *de.hypotheses.org* ist bisher mit wenigen Ausnahmen von geschichtswissenschaftlichen Blogs geprägt.

¹⁷ <http://sammelpunkt.philo.at:8080/cgi/latest> sowie <http://philpapers.org/>

¹⁸ <http://philsci-archive.pitt.edu>

¹⁹ <http://www.kritikon.de/> sowie <http://www.sicetnon.org>

²⁰ <http://www.romanistik.de> und <http://www.mediaevum.de/haupt2.htm>

²¹ <http://hsozkult.geschichte.hu-berlin.de/>

²² <http://web.fu-berlin.de/phin/>

²³ <http://www.gib.uni-tuebingen.de/image>

²⁴ <http://edoc.hu-berlin.de/histfor/>

²⁵ <http://www.uni-marburg.de/fb03/ivk/mjr>

Digitale Publikationsformen sowie die digitale Dissemination von Forschungsergebnissen und -projekten sind an der Schnittstelle zwischen digitaler Datenaufbereitung und -analyse anzusiedeln. Sieht man von einer immer größer werdenden Anzahl von E-Publikationen ab, die zumeist von Bibliotheken gehostet und den wissenschaftlichen Communities zur Verfügung gestellt werden, so ist erkennbar, dass gerade in den Geschichtswissenschaften neue Publikationsplattformen rasant an Bedeutung gewinnen. Hierzu zählen beispielsweise *Das Altägyptische Totenbuch. Ein digitales Textzeugenarchiv*²⁶, *EGO | Europäische Geschichte Online*²⁷, *Docupedia-Zeitgeschichte*²⁸ und die im Aufbau befindliche *International Encyclopedia of the First World War – 1914-1918 online*²⁹. All diesen Projekten ist gemein, dass hier nicht nur Forschungsergebnisse präsentiert werden, sondern den Nutzern zudem die Möglichkeit gegeben wird, Daten und Forschungsergebnisse nachzunutzen bzw. sie selbst ggf. mit weiteren Informationen anzureichern. Für die digitale Musikwissenschaft existieren folgende Informationsportale: *Virtuelle Fachbibliothek Musikwissenschaft (ViFA Musik)*³⁰, *muwimedial* (Stand 2010, wird nicht länger gepflegt)³¹ sowie die *WWW Sites of Interest to Musicologists der American Musicological Society (AMS)*³².

Elektronisches Publizieren in der Kunstgeschichte erfolgt bspw. im *Kunstgeschichte – Open Peer Reviewed Journal*³³, das insofern eine besondere Neuerung gegenüber anderen elektronischen Zeitschriften und Publikationsportalen darstellt, da es den Publikationsprozess nicht nur ins Digitale überträgt, sondern auch das traditionelle Review-Verfahren ablöst. Kommunikation über aktuelle Fachinformation erfolgt darüber hinaus in Form von Blogs – hier ist etwa der Blog auf *arthistoricum.net* zu nennen –, Mailinglisten oder Portalen wie dem *Portal Kunstgeschichte*.

Digitale Kommunikationsformen sind auch in der Philosophie kaum durchgehend etabliert, als ein Beispiel für neuere Experimente kann das Netzwerk *Philosophie der Lebenswissenschaften* in Deutschland gelten.³⁴ Prozess- und Ergebnisformate der Forschung (Tagungen, Workshops, Publikationen, Audio-Mitschnitte) werden daneben online rezipiert und vorgestellt – in Form von Internetauftritten, Blogs, Mailinglisten, und der H-Soz-u-Kult-ähnlichen *Information*

²⁶ <http://www.totenbuch.awk.nrw.de/>

²⁷ <http://www.ieg-ego.eu/>

²⁸ <http://docupedia.de/zg/Hauptseite>

²⁹ <http://www.1914-1918-online.net/>

³⁰ <http://www.vifamusik.de/>

³¹ <http://www.muwimedial.de/projects.php>

³² <http://www.ams-net.org/www-musicology.php>

³³ <http://www.kunstgeschichte-ejournal.net/>

³⁴ <http://www.philbio.de>

Philosophie.³⁵ Einzelpersonen, Projekte und Institute stellen Digitalisate im Fall begrenzter Empfängerkreise durch jeweils individuelle lokale Softwareplattformen zur Verfügung, die z.B. digitale Semesterapparate unterstützen. Vor allem aber erproben sie für weitere Disseminationswege neue, teils mit weiteren Daten angereicherte Vermittlungsformen wie Podcasts, z.B. durch die *philosophischen Audiotheken* der TU Darmstadt und der Universität Wien, kaum durch Lösungen wie iTunes.³⁶

³⁵ <http://www.information-philosophie.de/>

³⁶ <http://audiothek.philo.at/>; <http://www.audioreader.org/>

3. Verwendung von DH-Verfahren: Stand in den Fachdisziplinen

Bezogen auf die aktuell bereits im Report berücksichtigten geistes- und kulturwissenschaftlichen Fächer wird im folgenden Abschnitt vorgestellt, welche *DH-Verfahren* in diesen Fächern zurzeit genutzt und/oder intensiv erforscht werden. Als Belege zur aktuellen Nutzung solcher *Hot Spots* – im Text **fett** hervorgehoben – werden jeweils konkrete Projekte und Tools – im Text *kursiv* dargestellt – benannt.

3.1 Philologien

Ein zentraler Bereich digital gestützter Datenerfassung und -aufbereitung in den Philologien ist die wissenschaftliche **Textedition**, also die erschließende Wiedergabe von Manuskripten, gedruckten Dokumenten und anderen Schriftzeugnissen. Auf der Grundlage von digitalen Faksimiles umfasst sie insbesondere (1) Verfahren der **Transkription** des Wortlautes, der **Kodierung** von Textstrukturen sowie weiteren Merkmalen nach den Richtlinien der *Text Encoding Initiative*, (2) der kommentierenden, erschließenden, textkritischen und kontextualisierenden **Annotation** des Textes, sowie (3) der Aufbereitung des kodierten Textes für die gedruckte oder elektronische Präsentation. Große Textarchive wie die *Digitale Bibliothek* von *TextGrid*³⁷ oder das *Deutsche Textarchiv*³⁸ legen stärkeren Wert auf Metadaten, die **Verwendung von Normdaten** (u.a. für Personen und Orte) und eine einfachere Textauszeichnung, während Einzleditionen komplexere und vielschichtige Textauszeichnungen verwenden. Die für Transkription, Kodierung und Annotation eingesetzte Software umfasst insbesondere XML-Editoren, von denen *Oxygen* am weitesten verbreitet ist; den gesamten Produktionsprozess digitaler Editionen decken Anwendungen wie TUSTEP und *TextGrid* ab. Speziell für die Publikation wird unter anderem SADE³⁹ eingesetzt sowie häufig speziell auf die Bedürfnisse des jeweiligen Projekts angepasste, auf besonderen Standards (eXist Datenbank und Technologien wie XSLT, XQuery, JavaScript und CSS) basierte Eigenentwicklungen.

Die Textedition ist insbesondere in der Germanistik – aber auch in der Anglistik, deutlich weniger in der Romanistik – ein aktives Tätigkeitsfeld digitaler Literaturwissenschaft. Hier liegt auch der Schwerpunkt der Nutzung von digitalen Forschungsumgebungen wie *TextGrid*, wobei

³⁷ www.textgrid.de

³⁸ www.deutschestextarchiv.de

³⁹ <http://www.bbaw.de/telota/projekte/digitale-editionen/sade/sade-1>

mittlerweile auch linguistische Verfahren und Tools für andere Disziplinen (Geschichtswissenschaft, Kunstgeschichte, Musikwissenschaft etc.) eingebunden sind. Im Bereich der digitalen Editionsphilologie werden historisch-kritische und genetische Editionen erstellt und, ähnlich wie in der Geschichtswissenschaft, verschiedene Publikationsmodelle erprobt, insbesondere Hybrid-Editionen (print und online-Fassungen eines edierten Werks, mit teils unterschiedlichen Inhalten) und reine Online-Editionen. CD-ROM-basierte digitale Editionen werden kaum noch verfolgt. Zwei aktuell besonders stark beforschte Verfahren der Editionsphilologie sind die Dokumentation und Visualisierung von Varianten sowie die automatische Kollation und Stemma-Erstellung.

Der *Catalog of Scholarly Digital Editions*⁴⁰ führt rund 50 Editionen deutschsprachiger Materialien auf, die an Institutionen in Deutschland, Österreich und der Schweiz durchgeführt werden. Wichtige jüngere oder laufende Projekte sind zum Beispiel die *Digitale Faustedition*⁴¹ (Würzburg/Frankfurt), das *Jean-Paul-Portal*⁴² (Würzburg) sowie die *Heinrich-Heine-Edition*⁴³ (Düsseldorf/Trier); außerdem zu nennen ist die Edition *Berliner Intellektuelle*⁴⁴ (Humboldt Universität Berlin). Ebenfalls in dem erwähnten *Catalog* aufgeführte französisch-, italienisch- oder spanischsprachige Editionsprojekte sind nur in Ausnahmefällen in Deutschland angesiedelt, demnach kann in der Romanistik nicht von einer aktiven digitalen Editionsphilologie im deutschsprachigen Raum gesprochen werden – was auch auf die englischsprachige Philologie zutrifft. Im englischsprachigen Raum gibt es hingegen eine aktive digitale Editionsphilologie, hier können nur wenige Beispielprojekte genannt werden. Die entsprechende Projektwebseite der *Text Encoding Initiative* listet eine Vielzahl von Projekten⁴⁵ auf, die sich vor allem Texten aus der englischsprachigen Literatur widmen. Exemplarisch für eine Reihe von Projekten, die in größerem Umfang elektronische Texte unter bestimmten thematischen, zeitlichen oder genrespezifischen Kriterien bereitstellen, seien hier auch das *Women Writers Project* der Brown University⁴⁶ oder das *Victorian Women Writers Project*⁴⁷ genannt.

Als weiteres Feld der Datenerfassung und -aufbereitung kann der Aufbau von linguistisch sowie literaturwissenschaftlich nutzbaren **Corpora, lexikographischen Werken** sowie **Grammatiken**

⁴⁰ <http://www.digitale-edition.de>

⁴¹ <https://faustedition.uni-wuerzburg.de/dev/project/about>

⁴² <http://www.jean-paul-portal.uni-wuerzburg.de/>

⁴³ <http://www.akademienunion.de/projektbeschreibungen/hheine.htm>

⁴⁴ <http://141.20.126.175/berliner-intellektuelle/>

⁴⁵ <http://www.tei-c.org/Activities/Projects/>

⁴⁶ <http://www.wwp.brown.edu/>

⁴⁷ <http://webapp1.dlib.indiana.edu/vwwp/welcome.do>

angesehen werden. Im Bereich des Aufbaus linguistischer Corpora sind die Aktivitäten sehr zahlreich, so dass hier lediglich einige prominente Aktivitäten exemplarisch benannt werden sollen – hier aus dem Bereich der Anglistik. Der Prototyp eines großen, maschinell verarbeitbaren Corpus ist das British National Corpus (BNC)⁴⁸, das als repräsentatives Corpus für das britische Englisch der späten 1970er und vor allem der 1980er Jahre gilt und mit 100 Millionen laufender Wortformen nach wie vor eines der großen Corpora des Englischen ist. Weitere Aktivitäten zeigen beispielhaft das Spektrum der anglistischen Arbeiten im Bereich der historischen Corpora. Hier kann vor allem das *Innsbruck Computer Archive of Machine-Readable English Texts* (ICAMET)⁴⁹ genannt werden. Das wohl prominenteste **lexikographische Projekt** in der englischsprachigen Lexikographie ist jedoch die Digitalisierung des *Oxford English Dictionary*, das bereits 1857 durch die Philological Society of London initiiert und seit über 100 Jahren weiterentwickelt wird. Eine kleine Anzahl weiterer Projekte befasst sich mit der **Onlinebereitstellung von Grammatiken** und Materialien zur englischen Linguistik, wie etwa *HyperGrammar*⁵⁰ (University of Ottawa), *The Chemnitz Internet Grammar*⁵¹ (Universität Chemnitz) oder *The Virtual Linguistics Campus*⁵² (Universität Marburg). **Verfahren der computergestützten Informationsanalyse** unterstützen und erweitern einerseits etablierte literaturwissenschaftliche Verfahren, andererseits ermöglichen sie auch gänzlich neue Verfahren des interpretierenden Umgangs mit Texten.

Ein etabliertes, traditionelles literaturwissenschaftliches Verfahren ist das *close reading*, d.h. die sorgfältige, kleinteilige, detaillierte Lektüre und Analyse eines oft vom Umfang her relativ begrenzten Textes. Es hängt eng mit dem Verfahren der **Annotation** und der **Kohärenzbildung** zusammen: durch Annotationen werden Lektürebeobachtungen am Text dokumentiert, während Kohärenzbildung auf der vergleichenden Verbindung ähnlicher Textpassagen beruht. Alle drei Verfahren kommen in so unterschiedlichen methodischen Ansätzen wie dem Strukturalismus, der Narratologie, der Dekonstruktion oder dem New Historicism zur Geltung. Sie werden von bestimmten Tools unterstützt, beispielsweise von *CATMA*⁵³, das neben komplexen Suchanfragen an Texte auch die manuelle Markierung, Annotation und Kategorisierung kleinteiliger Textpassagen erlaubt, und zwar auch als kollaboratives Verfahren.

⁴⁸ British National Corpus: <http://www.natcorp.ox.ac.uk/>

⁴⁹ <http://www.uibk.ac.at/anglistik/projects/>

⁵⁰ <http://www.uottawa.ca/academic/arts/writcent/hypergrammar/grammar.html>

⁵¹ <http://www.tu-chemnitz.de/phil/english/InternetGrammar/frontend/index.html>

⁵² <http://linguistics.online.uni-marburg.de/>

⁵³ <http://www.catma.de/>

Weitere Verfahren, die ebenfalls die detaillierte Lektüre und Analyse von Texten unterstützen, führen zugleich zu neuen Formen des Lesens und der Kontextualisierung. Man spricht hier in umgekehrter Analogie zum *close reading* von *distant reading* – einem Begriff, den Franco Moretti mit *Graph, Maps, Trees. Abstract Models for a Literary History* (2005) geprägt hat. Derartige Verfahren des *distant reading* sind insbesondere dabei hilfreich, vergleichbare lexikalische oder stilistische Phänomene in einer umfangreicheren Textmenge zu ermitteln, indem sie mehr oder weniger komplexe Suchanfragen unterstützen. Hierunter fallen Verfahren wie die **Nutzung von Konkordanzfunktionen** und **Kollokationsanalysen**, aber auch die Suche nach bestimmten grammatikalischen Konstruktionen. Die ersten beiden Verfahren beruhen auf dem Prinzip der Volltextsuche und verlangen bei Vorliegen digitaler Fassungen der untersuchten Texte nur wenig Vorprozessierung. Für sie stehen zahlreiche, teils schon lange existierende Anwendungen zur Verfügung; hier seien als Beispiele Klassiker wie *MonoConc*⁵⁴, Mike Scotts *WordSmith Tools*⁵⁵ oder Lawrence Anthonys *AntConc*⁵⁶ genannt; in jüngerer Zeit sind auch web-basierte Services wie die *Voyant Tools*⁵⁷ (University of Alberta, Canada) verfügbar.

Die Suche nach grammatikalischen Konstruktionen erfordert die Aufbereitung der Texte mit Verfahren, die die Literaturwissenschaft von der Linguistik entlehnt. Insbesondere sind dies Verfahren der **Tokenisierung**, **Lemmatisierung** und des **Part-of-Speech-Tagging**, für die die Linguistik zahlreiche Tools zur Verfügung stellt. Part-of-Speech Tagging ermöglicht automatisiert und mit hoher Genauigkeit ($\geq 95\%$) die automatische Wortartenauszeichnung vorwiegend auf Basis probabilistischer Modelle. Beispiele hierfür sind der *TreeTagger*⁵⁸ (Universität Stuttgart) oder der *Stanford Part-of-Speech-Tagger*⁵⁹. Einige Part of Speech Tagger wie der *TreeTagger* ermöglichen eine **automatische Lemmatisierung**. Unter den automatischen und probabilistisch arbeitenden Annotationswerkzeugen werden vor allem syntaktische Parser, wie zum Beispiel der *Stanford Parser*, eingesetzt. Solche Werkzeuge existieren mittlerweile nicht nur für das Englische, sondern für eine Vielzahl weiterer Sprachen. Daneben existieren manuelle Werkzeuge zur Annotation linguistischer Merkmale, die sich nicht oder nur eingeschränkt automatisch erfassen lassen. Dazu zählen vor allem Merkmale der Diskursebene, wie zum Beispiel Mittel der lexikalischen Kohäsion oder spezifische

⁵⁴ <http://www.athel.com/mono.html>

⁵⁵ <http://www.lexically.net/wordsmith/>

⁵⁶ http://www.antlab.sci.waseda.ac.jp/antconc_index.html

⁵⁷ <http://voyant-tools.org/>

⁵⁸ <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>

⁵⁹ <http://nlp.stanford.edu/software/index.shtml>

lexikogrammatische Merkmale, die zum Teil theoriespezifisch beschrieben werden, wie zum Beispiel Verbklassifizierungen in der Systemisch Funktionalen Linguistik. Ein Werkzeug, das besonders zur manuellen Annotation von Kohäsionsmerkmalen eingesetzt wird, darüber hinaus aber auch für die Annotation anderer Merkmale auf mehreren Ebenen (die sogenannte multi-layer annotation) einsetzbar ist, ist *MMAX2*⁶⁰, ein Werkzeug zur manuellen Mehrebenenannotation.

Neben den hier benannten *stand-alone* Werkzeugen sind in jüngerer Zeit auch sogenannte **Frameworks** im Einsatz, die die **Erstellung von Verarbeitungsketten aus mehreren Annotations- und Analysewerkzeugen** sowie die **numerische Auswertung** und **Visualisierung** der Ergebnisse ermöglichen. Gut etabliert sind hier GATE⁶¹, MALLET⁶² sowie Tesla⁶³. Daneben haben sich Programmiersprachen mit spezifisch linguistischen Modulen wie beispielsweise das Python-basierte *Natural Language Toolkit* (NLTK)⁶⁴ und das Statistikpaket R etabliert. NLTK wird im Wesentlichen eingesetzt, um Corpora vorzuverarbeiten, zu annotieren und zu durchsuchen sowie um Verarbeitungsschritte zu serialisieren. R dient vor allem der Aufbereitung und statistischen Auswertung textueller Daten, es verfügt weiterhin über umfangreiche Bibliotheken zur Visualisierung von Daten auf der Basis statistischer Auswertungen. All diese Verfahren werden intensiv erforscht und für die Forschung genutzt, von einer Nutzung in der Breite der literaturwissenschaftlichen Forschungslandschaft und der Ausschöpfung ihres Potentials für philologische Forschung kann aber nicht gesprochen werden. Einen noch größeren Teil des Forschungsprozesses als die erwähnten Frameworks decken **virtuelle Forschungsumgebungen** wie *TextGrid* ab (das von Literatur- aber auch von Musikwissenschaftlern und Historikern und anderen genutzt wird), die von der Digitalisierung über die modulare Einbindung von Tools bis hin zu kollaborativen Prozessen und Publikationstechniken alle Prozesse und Verfahren im DH-Kontext unterstützen.

Zwei genuin digitale Verfahren der Literaturwissenschaft, die sich von vergleichbaren nicht-digitalen Verfahren so stark unterscheiden, dass sie als neue *DH-Verfahren* bezeichnet werden können, sind die **Stylometrie** und das **Topic Modeling**, die beide zum Bereich des Text Mining gehören. Ersteres ist das digitale Pendant zur Stilistik, d.h. die digital gestützte Erhebung von wiederkehrenden, quantifizierbaren stilistischen Phänomenen in Texten. Auf der Grundlage der

⁶⁰ <http://mmax2.sourceforge.net/>; außerdem: <http://www.wagsoft.com/CorpusTool/>,
<http://www.wagsoft.com/RSTTool/>.

⁶¹ <http://gate.ac.uk/>

⁶² <http://mallet.cs.umass.edu/>

⁶³ <http://www.spinfo.phil-fak.uni-koeln.de/spinfo-forschung-tesla.html>

⁶⁴ <http://nltk.org/>

Erkennung solcher Muster in großen Mengen von Texten wird das halbautomatische Vergleichen und Klassifizieren ermöglicht; unter anderem können hierdurch anonyme Texte einem bekannten Autor zugeschrieben werden. Solche Verfahren erlauben eine neue Sicht auf umfangreiche Textsammlungen und verwenden hierfür Algorithmen aus der Statistik; als Tool hervorzuheben sind die Skripte *stylo* und *classify* für die Statistikorientierte Programmiersprache R.⁶⁵ Das digitale Pendant zur thematischen Lektüre ist dagegen das Topic Modeling, das eine automatische Extraktion von Themengebieten aus einer umfangreichen Sammlung von Texten meint. Hier stehen die Gruppierung von Texten nach dominanten Themen sowie die Erhebung von thematischen Trends über die Zeit hinweg im Zentrum des Erkenntnisinteresses, d.h. literaturhistorische Fragestellungen mit Relevanz auch für die Ideengeschichte. Beide Verfahren, die Stylometrie und das Topic Modeling, werden von einzelnen Forschern erprobt und teils in Workshops unterrichtet, können in den Literaturwissenschaften in Deutschland jedoch nicht als etabliert gelten.

Über eine reine Analyse hinaus geht die **Visualisierung** von Ergebnissen, die mit quantitativen Verfahren der Textanalyse (wie die oben beschriebenen) erhoben wurden. Hier kann ein eher illustrativer von einem heuristischen bzw. analytischen Einsatz dieses Verfahrens unterschieden werden. Im letzteren Fall ist die Visualisierung selbst ein Mittel der Erkenntnis und fördert durch die Muster, die an ihr sichtbar werden, auch neue Fragestellungen zu Tage. Als Beispiel kann das Projekt *eTraces*⁶⁶ genannt werden (Universität Leipzig, Göttingen Centre for Digital Humanities, Leibniz Institute for Social Sciences Bonn), in dem intertextuelle Beziehungen visualisiert werden sollen. Visualisierungen kommen in teils einfachen Linien- oder Balkendiagrammen vor (dann werden meist Frequenzen oder Frequenzverteilungen visualisiert, wie beispielsweise in *Voyant* oder *Antconc*), teils werden komplexere Verhältnisse visualisiert (wie in den Cluster Analysis Graphen und dem Consensus Tree der oben erwähnten stylometrischen Skripte). Oder es werden Netzwerke von miteinander verbundenen Einzeldaten visualisiert, was unter anderem mit den Tools *Gephi*⁶⁷ oder *Pajek*⁶⁸ möglich ist. Während die ersten beiden Formen der Visualisierung durchaus von einigen Literaturwissenschaftlern praktiziert werden, scheint die **Netzwerkanalyse** eher Historiker als Literaturwissenschaftler anzusprechen. Ein mögliches Einsatzgebiet solcher Instrumente für die Literaturwissenschaft ist die Analyse von Briefnetzwerken. Die Digitalisierung, Aufbereitung und Analyse von großen

⁶⁵ <https://sites.google.com/site/computationalstylistics/>

⁶⁶ <http://etraces.e-humanities.net/>

⁶⁷ <http://gephi.org/>

⁶⁸ <http://pajek.imfm.si/doku.php>

Briefkorpora sowie der enthaltenen Netzwerkstrukturen wird derzeit in mehreren Projekten verfolgt, wie z.B. in dem Projekt *Briefe und Texte aus dem intellektuellen Berlin um 1800*⁶⁹.

3.2 Geschichtswissenschaft

Die Geschichtswissenschaft gehört zu den geisteswissenschaftlichen Disziplinen, die über keine geschlossenen bzw. klar begrenzten Forschungsfragen und Forschungsgegenstände verfügen. Das Erkenntnisinteresse der Historiker reicht oftmals in benachbarte Disziplinen und deren Themen hinein, so dass vielfältige Methoden und Verfahren benutzt und verwendet werden. Diese wissenschaftliche und zum Teil sehr stark interdisziplinär ausgerichtete Vielfalt des Faches lässt sich im Folgenden nur partiell abbilden – dies gilt gerade für die historischen Hilfswissenschaften –, so dass sich auf die zentralen Elemente der in der Geschichtswissenschaft verwendeten DH-Verfahren zu konzentrieren gilt.

Verfahren der digitalen Datenerfassung und -aufbereitung finden sich in bestimmten Kernbereichen. Über die reine Digitalisierung von Quellenmaterial und gedruckten bzw. publizierten zeitgenössischen Werken hinaus werden mit dem Ziel, **wissenschaftliche historisch-kritische Editionen** zu erstellen, Schriftzeugnisse auf Zeichenebene transkribiert, annotiert, Aussagen des Textes historisch-kritisch kommentiert, oftmals der Entstehungsverlauf schriftlicher Zeugnisse miteinander verglichen und die edierten Texte mit Metadaten, biblio- und biographischen Informationen und ggf. Normdaten auf Textebene angereichert. Bei einer Vielzahl der derzeitigen Editionsprojekte⁷⁰ – hierbei werden in der Geschichtswissenschaft im Prinzip die gleichen *DH-Verfahren* wie in den Philologien eingesetzt⁷¹ – geschieht dies – ebenfalls wie in den Philologien – entlang der Richtlinien der *Text Encoding Initiative* (TEI) mit dem Ziel, diese Forschungsergebnisse digital bzw. gedruckt zu publizieren. Gerade bei digitalen Editionsprojekten ist ein hoher Entwicklungsbedarf erkennbar, da Fragen der Archivierung, der Distribution und der Langzeitverfügbarkeit zur Nachnutzung des edierten Quellenmaterials derzeit nur von wenigen Projekten behandelt werden. An dieser Stelle sind exemplarisch die

⁶⁹ <http://tei.ibi.hu-berlin.de/berliner-intellektuelle/>

⁷⁰ Z.B.: *Controversia et Confessio* - <http://controversia-et-confessio.adwmainz.de/index.php?id=600>; Editionsprojekt *Leichenpredigten* - <http://www.personalschriften.de/> ; *Carl-Maria-von-Weber-Gesamtausgabe* - <http://www.weber-gesamtausgabe.de/de/Index>; *Virtuelles Skriptorium St. Matthias* - <http://stmatthias.uni-trier.de/>; *Eugenio Pacelli - Kritische Online-Edition der Nuntiaturreportagen von 1917-1929* - <http://www.pacelli-edition.de/>; *Alfred Escher-Briefedition* - <http://www.briefedition.alfred-escher.ch/>; *Mitteldeutsche Selbstzeugnisse der Zeit des Dreißigjährigen Krieges* - <http://www.mdsz.thulb.uni-jena.de/sz/index.php>

⁷¹ Die hierzu verwendeten XML-Editoren bzw. Softwarelösungen sind die selben, die bereits im Kapitel 4.1 Philologien beschrieben wurden.

*Digitale Bibliothek von TextGrid*⁷², *Monasterium*⁷³ und das Projekt *Archiv-, Editions- und Distributionsplattform für Werke der Frühen Neuzeit* (AEDit Frühe Neuzeit)⁷⁴ zu nennen.

Digitale Datenerfassung und -aufbereitung der historischen Disziplin erfolgt auch unter Nutzung von **Virtuellen Forschungsumgebungen** (VRE). Hierbei ist zwischen zwei unterschiedlichen Verfahren zu differenzieren. Während sich z.B. mit GAMS⁷⁵ Textdokumente mit X-Technologien erfassen und erschließen lassen, so dass sie maschinenlesbar vorliegen, können mit FuD⁷⁶ – einem modular aufgebauten Datenbanksystem und einem eher objektbezogenen System – beliebige Forschungsdaten und Datenobjekttypen erfasst, mit Metadaten angereichert und z.B. annotiert bzw. für eine Publikation vorbereitet werden. Gerade dieser zuletzt genannte VRE-Typ besitzt einen eher generischen Charakter, da durch ihn eine größere Bandbreite an Quellen- und Dokumenttypen erfasst und aufbereitet werden kann. Dadurch lassen sich die vielfältigen historischen Forschungsmethoden und Workflows, in denen höchst unterschiedliche Quellentypen analysiert werden, näher abbilden. Sie werden begleitet von eher material- oder fachspezifischen VREs wie MOM (Monasterium)⁷⁷, die aber ebenfalls Lösungen entwickeln, die verallgemeinert werden können. Die **Entwicklung und Realisierung Virtueller Forschungsumgebungen** ist einer der *Hot Spots* in der Geschichtswissenschaft, was sich u.a. auch daran ablesen lässt, dass im vergangenen Jahr die Arbeitsgemeinschaft *Virtuelle Forschungsumgebungen in der Geschichtswissenschaft* gegründet wurde⁷⁸, die einen permanenten Erfahrungsaustausch auf diesem elaborierten Feld ermöglicht. In diesen Kernbereich der digitalen Datenerfassung und -aufbereitung fällt aber auch eine Vielzahl kommerzieller **Literaturverwaltungsprogramme**, die von Wissenschaftlern nicht nur zur bibliographischen Erschließung von Forschungsliteratur verwendet werden, sondern auch für die **Aufbereitung von Quellen- und Archivbeständen**.⁷⁹ In diesem Kontext der basalen DH-Verfahrenstechniken ist auch der Bereich des **kollaborativen Schreibens** zu sehen, der in den vergangenen Jahren auch in den Geschichtswissenschaften massiv an Bedeutung gewann und

⁷² <http://www.textgrid.de/>

⁷³ <http://www.monasterium.net/>

⁷⁴ <http://www.hab.de/forschung/projekte/aedit.htm>

⁷⁵ Geisteswissenschaftliches Asset Management System (GAMS): <http://gams.uni-graz.at/>

⁷⁶ Forschungsnetzwerk und Datenbanksystem (FuD): <http://fud.uni-trier.de/>

⁷⁷ <http://www.monasterium.net/> bzw. als Arbeitsumgebung <http://www.mom-ca.uni-koeln.de>

⁷⁸ U.a. von Vertretern von H-Soz-Kult, des Georg-Eckert-Instituts für Internationale Schulbuchforschung, der Bayerischen Staatsbibliothek, der Max-Weber-Stiftung, der Historisch-Kulturwissenschaftlichen Informationsverarbeitung (Köln), des Trier Center for Digital Humanities und des Leibniz-Instituts für Europäische Geschichte.

⁷⁹ Hier wären eine Vielzahl an Beispielen für Literaturverwaltungs- und Wissensmanagementprogrammen zu nennen, exemplarisch für die historisch arbeitenden Geisteswissenschaften sind CITAVI - <http://www.citavi.com/de/>; Endnote - <http://endnote.com/>; zotero - <http://www.zotero.org/>.

der vor allem dadurch ermöglicht wurde, dass neue Tools und Werkzeuge entstanden, mit deren Hilfe Autoren gemeinsam zeitgleich und ortsungebunden Texte erstellen können.⁸⁰

Sämtlichen Forschungsprojekten in der Geschichtswissenschaft, die sich der Hilfe von DH-Verfahren bedienen, ist dabei gemein, dass sie verstärkt ihre Forschungsdaten mit **Normdaten** anreichern, so z.B. aus der *Gemeinsamen Normdatei* der Deutschen Nationalbibliothek (GND)⁸¹, den *Getty Vocabularies*⁸² oder dem *Virtual International Authority File (VIAF)*⁸³. Durch die Anwendung dieser Standards können Forschungsdaten unterschiedlichster Provenienz miteinander verknüpft werden. Zugleich wird hierdurch eine grundsätzliche Basis für **Linked Open Data (LOD)** geschaffen. Hierbei handelt es sich um keine spezifische Entwicklung der Geschichtswissenschaft, sondern vielmehr um eine, die in zahlreichen geisteswissenschaftlichen Fachdisziplinen vorzufinden ist (Siehe z.B. Abschnitt 3.1 – Philologien).

Im Bereich der Quellenanalyse und Quellenaufbereitung sind folgende *DH-Verfahren*, die zum Teil bereits seit Jahrzehnten angewandt werden und deshalb als etabliert gelten können, zu nennen. Ein Beispiel sind **statistisch quantitative Analyseverfahren**⁸⁴. Diese werden verwendet, um Datenmaterial, wie z.B. Statistiken, Einkommensverteilungen, Budgetberechnungen, demographische Informationen, mit Hilfe von meist aus den Sozialwissenschaften entlehnten Software-Tools empirisch zu analysieren. Ein weiteres Beispiel sind **Digitale Kartenserver und -atlanten** (wie *IEG Maps* und *AtlasEuropa*⁸⁵), vor allem aber **historische Geoinformationssysteme**⁸⁶. Diese ermöglichen es, dynamische Zeit-Raum-Objektbeziehungen kartographisch zu modellieren (wie *HGIS Germany*⁸⁷ oder das britische GBHGIS⁸⁸).

Neben dem **Encoding** – hier bieten sich der Geschichtswissenschaft ähnliche Nutzungsszenarien und -möglichkeiten wie den anderen geistes- und kulturwissenschaftlichen Disziplinen – spielen **Netzwerkanalyse** und auch **Visualisierungstechniken & 3D-**

⁸⁰ Neben google.docs und etherpad ist auch die VRE *edumeres.net* zu nennen, die kollaboratives Schreiben und gemeinsames Publizieren ermöglicht, siehe: <http://www.edumeres.net/> [Edumeres ist eine interdisziplinär konzipierte VRE, die den verschiedenen an der Bildungsmedienforschung beteiligten Disziplinen eine gemeinsame Arbeitsumgebung bietet].

⁸¹ http://www.dnb.de/DE/Standardisierung/Normdaten/GND/gnd_node.html

⁸² <http://www.getty.edu/research/tools/vocabularies/>

⁸³ <http://viaf.org/>

⁸⁴ Z.B. <http://www.histat.gesis.org/>; als Softwareumgebung wird das in den Sozialwissenschaften gebräuchliche SPSS verwendet.

⁸⁵ <http://www.atlas-europa.de/>

⁸⁶ Überblick: <http://www.hgis.org.uk/>

⁸⁷ <http://www.hgis-germany.de/>

⁸⁸ <http://www.gbhgis.org/>

Modellierung/-Rekonstruktion als DH-Verfahren für die Geschichtswissenschaften eine Rolle. Visualisierungstechniken und die 3D-Modellierung sind überwiegend aus den Disziplinen Architektur und Archäologie entlehnt. Mit deren Hilfe werden beispielsweise Stadt- und architekturhistorische Fragestellungen behandelt, um so Differenzen zwischen Plänen und Entwürfen und den tatsächlichen späteren Realisierungen graphisch darstellen zu können; oftmals werden diese mit GIS-Verfahren gekoppelt eingesetzt. Netzwerkanalysen stellen ein weiteres in der Geschichtswissenschaft eingesetztes DH-Verfahren dar, in denen die historischen Grundkonstanten Raum, Zeit, Akteure, Sprache und Inhalte nicht nur zueinander in Beziehung gesetzt werden, sondern erst durch die Anwendung neuartiger *DH-Verfahren* große Daten- und Quellenmengen überhaupt analysierbar werden. Netzwerkanalysen können in drei Dimensionen unterteilt werden. 1. Sozialstatistische und prosopographische Netzwerkanalysen, die Beziehungsgeflechte zwischen Personen und Gruppen darstellen;⁸⁹ 2. Netzwerkanalysen, die akteurszentriert Vernetzungen zeit- und raumbezogen aufzeigen; sowie 3. Diskurs-Netzwerkanalysen, deren Fokus auf der Untersuchung von Diskursen in Personen-Raum-Zeit-Geflechten liegt. Als state-of-the-art können hierfür exemplarisch folgende Projekte genannt werden: *Mapping the Republic of Letters*⁹⁰ und das *Briefnetzwerk. Visualisierung von mehrdimensionalen Informationsstrukturen in Briefcorpora*⁹¹.

3.3 Kunstgeschichte

Die Erfassung, Aufbereitung, Analyse und Auswertung digitalisierter Bilder sowie das Anfertigen virtueller Modelle stehen dem Gegenstandsbereich der Kunstgeschichte entsprechend auch im *DH-Bereich*, der Digitalen Kunstgeschichte, im Mittelpunkt. Dies bedeutet allerdings nicht, dass auch digitale Verfahren, die schwerpunktmäßig in anderen Disziplinen entwickelt wurden, wie etwa die **Visualisierung von Netzwerken** (z.B. Darstellung von Kontaktverhältnissen, Schulbildung) oder die **Darstellung geographischer Informationen auf einer Karte** (z.B. Provenienz), in der Digitalen Kunstgeschichte nicht zur Anwendung kommen. Allerdings soll an dieser Stelle auf die der Digitalen Kunstgeschichte inhärenten Verfahren fokussiert werden, um deren Potential für die kunsthistorische Forschung darzulegen.

⁸⁹ Hierfür können beispielsweise Pajek, UCINET, Gephi oder auch Cytoscape – die Liste ließe sich beliebig erweitern - verwendet werden. Für eine ausführliche Diskussion über die Verwendung von Netzwerk-Visualisierung- und Analysetools siehe: <http://digitalhumanities.org/answers/topic/what-is-the-best-software-package-for-social-network-analysis>

⁹⁰ <http://republicofletters.stanford.edu/> Es bleibt bei diesem Projekt zu hoffen, dass zukünftig auch die Inhalte der Korrespondenz über das Web-UI analysierbar sein werden.

⁹¹ <http://kompetenzzentrum.uni-trier.de/de/projekte/projekte/briefnetzwerk/>

Als zwei bedeutende fachspezifische *DH-Verfahren* zur Datenaufbereitung und -auswertung sind die **crossmediale semantische Verknüpfung** und das **Social Image Tagging** mit dem Spezialbereich der so genannten **Gamification**. Ausgangspunkt crossmedialer semantischer Verknüpfung ist die durch Digitalisierung entstandene Möglichkeit, Informationen miteinander zu verknüpfen. Ebenso wie textuelle Informationen sich semantisch verknüpfen lassen, können auch visuelle Informationen semantisch miteinander oder auch mit anderen Informationseinheiten wie Ortsdaten, Quellenbeständen etc. verknüpft werden, um sie für eine wissenschaftliche Nutzung fruchtbar zu machen. Als einschlägiges Projekt im Bereich der Kunstgeschichte, aber auch der Archäologie, ist hier *HyperImage*⁹² zu nennen, das eine Fortführung als *Meta-Image* zur Einbindung des in *HyperImage* entwickelten Bildannotationstools in die Bilddatenbank *prometheus* erfuhrt. Das Social Image Tagging, also die Erschließung und Auswertung einer großen Menge von Bilddaten unter Einbeziehung von Internetnutzern, stellt ein besonders reizvolles, aber auch kontrovers diskutiertes Feld dar, wird hierbei doch weniger auf Expertenwissen als auf das Wissen der Massen gesetzt. Erfolgreich eingesetzt wird dieses Verfahren derzeit z.B. in dem Projekt *Your Paintings Tagger*⁹³, bei dem Internetnutzer den kompletten Bestand an Ölgemälden öffentlicher Sammlungen von Großbritannien mit Annotationen versehen. Social Image Tagging kann darüber hinaus mit einer Komponente des Spiels versehen werden: Die User befinden sich in einer Art Wettkampfsituation, in der es gilt, die treffendsten Schlagworte zu vergeben, so dass die zu erwartende Qualität der Ergebnisse steigt. Eine solche Gamification des Taggingprozesses wird zurzeit beispielsweise im Projekt *Artigo* erprobt.⁹⁴

Als Hot Spot im Bereich der Datenerfassung und -analyse ist die **Automatic Image Recognition** zu nennen, die als ein Verfahren bezeichnet werden kann, das genuin digital ist. In der automatischen Bild- bzw. Mustererkennung liegt ein sehr großes Potential für die Kunstgeschichte, wie es z.B. das Forschungsprojekt *What Makes Paris Look Like Paris?*⁹⁵ (Carnegie Mellon University, Inria, Ecole Normale Supérieure Paris) eindrücklich demonstriert, sowie für andere mit Bildern arbeitende Disziplinen wie etwa der Archäologie, denn hierdurch kann direkt und ohne zeitaufwändigen Umweg über Annotationen mit Bildinformationen gearbeitet werden. In Analogie zu den insbesondere in den Literaturwissenschaften vorangetriebenen *DH-Verfahren* **Stylometrie** und **Topic Modeling** (vgl. Abschnitt Philologien)

⁹² <http://www2.leuphana.de/meta-image/index.html>

⁹³ <http://tagger.thepcf.org.uk/>

⁹⁴ <http://www.artigo.org/>

⁹⁵ <http://graphics.cs.cmu.edu/projects/whatMakesParis/> (Projektwebsite),
<http://www.zeit.de/digital/internet/2012-08/street-view-fingerabdruck> (ZEIT-Artikel)

können so etwa Handschriften eines Künstlers automatisch erkannt oder zeitliche Entwicklungen bestimmter kunsthistorischer Motive verfolgt werden. Der jeweiligen Forschungsfrage gemäß kann eine Mustererkennung dabei unterschiedlich aufwändig sein und mittels verschiedener Erkennungsalgorithmen sowie mit Unterstützung maschinellen Lernens arbeiten. Grob lässt sich aber zwischen dem Erkennen einfacher Muster wie Rechtecken, Dreiecken oder Polygonen mit überschaubar vielen Punkten und dem Erkennen komplexer Form- und Farbzusammenhänge unterscheiden. So ist z.B. die Erkennung von Händen sehr aufwändig, da sie nicht immer identisch dargestellt sind. Hände können z.B. gefaltet, zur Faust geballt oder mit gespreizten Fingern dargestellt sein. Hier gibt es nicht nur ein Muster, nach dem gesucht werden muss, sondern gegebenenfalls tausende von Mustern. Es bedarf hier gegenüber dem einfacheren Fall wesentlich komplexerer Erkennungsalgorithmen. Als prominentes Projekt im deutschsprachigen Raum gilt das Projekt *Computerassisted Detection and Analysis of Medieval Legal Gestures* der Forschungsgruppe *Heidelberg Collaboratory for Image Processing*, in dem Gesten in mittelalterlichen Handschriften automatisch ausgewertet wurden.⁹⁶

Weiterhin hat sich die **3D-Modellierung** als Verfahren zur Darstellung und Erfassung komplexer baugeschichtlicher Sachverhalte etabliert. Im Rahmen des europäischen Projekts *3D-Bridge - Transferring Cultural Heritage with New Technology*⁹⁷ wurde z.B. Filippo Juvarras originärer Entwurf zum Concorso Clementino, einem herrschaftlichen Palazzo mit Garten, auf Grundlage seiner Architekturzeichnungen in Form eines begehbaren 3D-Modells realisiert.⁹⁸ Ein Forschungsschwerpunkt zu digitalen Architekturmodellen im Zusammenhang mit der Residenzenforschung konnte außerdem in den letzten Jahren an der Professur für Bayerische Kunstgeschichte an der Ludwig-Maximilians-Universität München entwickelt werden. Das Besondere an der Erstellung virtueller Modelle – wie auch aller anderen Ergebnisse von Visualisierungsprozessen – ist hierbei ihr Zug, ihr Ergebnisformat selbst wieder zum Gegenstand neuerlicher Reflektion anzubieten. Derzeit besteht allerdings noch ein Forschungsdesiderat in Bezug auf die Erarbeitung einer Genealogie, Typologie, Theorie und Kritik virtueller Modelle.

In starker Differenz gegenüber den bisher vorgestellten Verfahren, im Grenzbereich von Kunstgeschichte und Wahrnehmungstheorie bewegend, ist das so genannte **Eye-Tracking** zu sehen. Eye-Tracking stellt ein Verfahren dar, das die Augenbewegung des Betrachters exakt zu

⁹⁶ <http://hci.iwr.uni-heidelberg.de/COMPVIS/research/gestures/>

⁹⁷ <http://www.arthis.jyu.fi/bridge/>

⁹⁸ <http://wissensgeschichte.biblertz.it/3d-bridge-html/index3D.html>

erfassen versucht, um damit primäre Vorgänge der Rezeption zu rekonstruieren. Die Ergebnisse der Rezeptionsforschung stellen dabei innerhalb der Kunstgeschichte ein wichtiges Korrektiv zur Rezeptionsästhetik dar, die wiederum innerbildliche Mittel als Rezeptionsvorgaben untersucht.

Kunsthistorische Forschung stellt eine Seite der Medaille im Fach Kunstgeschichte dar. Die andere Seite ist immer auch die der Kunstvermittlung, wie sie institutionalisiert in Museen betrieben wird. Die **Vermittlung von Kunst mittels mobiler Geräte** wie Smartphones stellt hierbei dann ein innovatives Gebiet dar, wenn solche Lernspiele der Gefahr einer zu starken Trivialisierung des Kunstbegriffs widerstehen. Innovative Ansätze zum musealen Einsatz von Smartphone-Apps, die QR-Code auslesen können, finden sich in dem Projekt *QRator*, das weniger auf die Vermittlung von Inhalten, als auf die dokumentierte Meinungsabgabe der Museumsbesucher zielt, so dass Besucher gleichsam zu einem Co-Kuratoren werden können, mit Möglichkeiten der Partizipation an einer über den einzelnen Ausstellungsbesuch hinausgehenden Diskussion über die jeweiligen Exponate.⁹⁹

Im deutschsprachigen Raum hat sich der Arbeitskreis *Digitale Kunstgeschichte* formiert¹⁰⁰, um die Diskussion über methodologische Fragen anzuregen und zu begleiten, um fachspezifische Perspektiven in den Digital Humanities zu entwickeln, einschlägige Projekte ideell und organisatorisch zu unterstützen und den Nachwuchs zu fördern.¹⁰¹ Außerhalb Deutschlands und in den USA ist insbesondere die *Software Studies Initiative* unter Leitung von Lev Manovich als treibende Kraft im Feld der Digitalen Kunst- sowie Kulturgeschichte zu nennen,¹⁰² die z.B. das äußerst hilfreiche Visualisierungstool *ImagePlot* entwickelt hat, mit dem sich alle Bilder einer Sammlung z.B. nach Datum, aber auch nach ihrer visuellen Beschaffenheit geordnet anzeigen lassen.¹⁰³

3.4 Archäologie

Die Archäologie hat von technischen Neuerungen insbesondere auf dem Gebiet der Vermessung und der Fotografie profitiert. Digitale Niveliergeräte/Theodolite und Kameras mit GPS gehören heute als Arbeitsgeräte zur Ausstattung der Grabungstechnik. Digitale Methoden

⁹⁹ <http://www.qrator.org/>

¹⁰⁰ Dessen konstituierende Tagung im Februar 2012 wurde von Stephan Hoppe, Hubertus Kohle, Katja Kwastek und Georg Schelbert organisiert.

¹⁰¹ <http://www.digitale-kunstgeschichte.de/wiki/Hauptseite/> [Der Arbeitskreis *Digitale Kunstgeschichte* ist aus diesem Grund als ein wichtiger Inputgeber für den vorliegenden Report zu nennen.]

¹⁰² <http://lab.softwarestudies.com/>

¹⁰³ <http://lab.softwarestudies.com/p/imageplot.html>

zur Datenerfassung in der Archäologie sind beispielsweise die **geoelektrische** und **geomagnetische Prospektion**, der Einsatz von **Bodenradar**, die **Fotogrammetrie** sowie das **3D-Scanning** von Bauwerken und Objekten.

Geoelektrische und **geomagnetische Prospektionen** dienen dazu, durch die Erfassung von Unterschieden in der elektrischen Leitfähigkeit des Bodens oder im Erdmagnetfeld unbekannte bzw. solche Strukturen im Boden (z.B. Gräber, Häuser) lagegerecht darzustellen, die nicht z.B. durch Luftbilder oder Laser-Impulse erfasst werden können. **Bodenradar** wird dabei ebenfalls zur Datenerfassung eingesetzt. Anwendungsbeispiele und Best-Practice-Richtlinien werden hierfür im EU-Projekt *ArcLand – Archeo Landscapes Europe*¹⁰⁴ genannt bzw. erstellt, an dem auch das Deutsche Archäologische Institut¹⁰⁵ beteiligt ist. Das Landesamt für Denkmalpflege und Archäologie Sachsen-Anhalt setzt ebenfalls solche Verfahren ein. **Fotogrammetrie** – differenzierbar in **terrestrische Fotogrammetrie** und **Aerofotogrammetrie** (Luftbildarchäologie) – ist vereinfacht ein Verfahren zur technischen Bildvermessung um Formen, Größen und Lage von Strukturen zu bestimmen und daraus Pläne und Karten zu erstellen. Durch das **3D-Scanning von Bauwerken und Objekten** mittels Laser-Technologie, das sich wohlgermerkt von einer 3D-Rekonstruktion unterscheidet, wird die Gewinnung dreidimensionaler raumbezogener Daten für eine steingenaue und photorealistische Dokumentation der archäologischen Befunde im Innen- und Außenbereich ermöglicht – bspw. durchgeführt vom Landesamt für Archäologie Sachsen.¹⁰⁶

Im Bereich der Datenaufbereitung in der Archäologie ist zum einen das Verfahren der **3D-Visualisierung** zu nennen, dabei unterscheidet man zwischen Bestandsaufnahme, die tatsächlich nur den Befund in situ wiedergibt, somit noch keine "Interpretation" ist, und der Rekonstruktion, in der virtuell Ergänzungen vorgenommen werden. Ziel einer 3D-Rekonstruktion könnte beispielsweise die Visualisierung des Raumbezugs einer Architektur sein - die Blickachsen bei einer domus oder die strategische Lage einer Befestigungsanlage. Digitale Auswertungsprozesse erfolgen zudem auch mittels **Bildannotation** sowie perspektivisch mit **Mustererkennung** (vgl. Abschnitt 3.3, Kunstgeschichte). Ein Beispiel für eine Plattform in der über verschiedene Sammlungen speziell zur Klassischen Archäologie nach Objekten gesucht werden kann, ist CLAROSnet.org (CLassical Art Research Online Research Services). CLAROS gewährleistet Interoperabilität der Daten durch die Nutzung von CIDOC-CRM und

¹⁰⁴ <http://www.archaeolandscapes.eu/>

¹⁰⁵ <http://www.dainst.org/de/>

¹⁰⁶ <http://www.archaeologie.sachsen.de/951.htm>

experimentiert mit einem Prototypen zur Bilderkennung, um die Suchfunktionalitäten zu erweitern.

3.5 Musikwissenschaften¹⁰⁷

Die digitale Musikwissenschaft zeichnet sich dadurch aus, dass ihre Methoden und Verfahren sehr stark interdisziplinär geprägt sind, so dass sich viele Überschneidungen mit verwandten oder angrenzenden Wissenschaften wie der Philologie, Soziologie, Ästhetik, Psychologie oder Akustik ergeben. Die folgende Darstellung nimmt diese Bereiche aus und beschränkt sich im Wesentlichen auf den Kernbereich der „Musik“ selbst und erhebt selbstverständlich keinerlei Anspruch auf Vollständigkeit.

Datenerfassung von Musik (im engeren Sinne) kann auf drei Ebenen erfolgen: Auf der **graphischen Ebene** sind neben Faksimiles proprietäre Tools zur Noteneingabe (*Finale*¹⁰⁸, *Sibelius*¹⁰⁹, *Score*¹¹⁰ usw.) sowie open-source-Tools zur Dateneingabe (*Lilypond*¹¹¹) zu nennen. Auf der **logischen Ebene** ist einzuordnen: der Noteneditor *Meise*¹¹², Datenformate wie MuseData, MusicXML, MEI¹¹³ (Music Encoding Initiative) usw. sowie die Erfassung musikalischer Katalogdaten mit *MerMEId* (Metadata Editor and Repository for MEI Data). Datenerfassung auf der **akustischen Ebene** findet bspw. in MIDI oder MP3 statt. Die Tools, die auf der graphischen Ebene benutzt werden, lassen sich teilweise auch zur Erzeugung/Erfassung von Daten auf der logischen und der akustischen Ebene nutzen. Mit Standard-Techniken (z. B. XSL-Transformationen) lassen sich Daten so konvertieren, dass sie teilweise in anderen Kontexten weitergenutzt werden können. Neben der Datenerfassung von Musik sind selbstverständlich in großem Ausmaß auch Texte zu erfassen. Hierfür werden reine Textverarbeitungsprogramme, OCR-Software sowie XML-Editoren (vorzugsweise *Oxygen*) genutzt, derzeit vor allem im editionsphilologischen Bereich.

Als *DH-Verfahren* der Datenaufbereitung können aufgeführt werden: die **Kollationierung** von Daten mit Hilfe der Ediom (auf der graphischen Ebene), das **Rendern von Musikcodierungen** mit SVG, z.B. zur Anzeige von MEI-Files und das **Audio-Text-Alignement** (akustische und

¹⁰⁷ Das Kapitel zu DH-Verfahren in den Musikwissenschaften wurde maßgeblich vom Musikwissenschaftlichen Seminar Detmold/Paderborn erstellt.

¹⁰⁸ <http://www.finalemusic.com/default.aspx>

¹⁰⁹ <http://www.sibelius.at>

¹¹⁰ <http://scoremus.com/>

¹¹¹ <http://lilypond.org/>

¹¹² <https://dev2.dariah.eu/wiki/pages/viewpage.action?pageId=7439804>

¹¹³ <http://music-encoding.org/>

graphische oder logische Ebene). Nicht Musik-bezogene Daten werden mit den Standardmethoden der jeweiligen verwandten Disziplinen bearbeitet.

Datenauswertungs-Prozesse erfolgen mithilfe von Tools wie *MIREX* (Music Information Retrieval Evaluation eXchange), das Funktionen wie **Melodiesuche**, **Ähnlichkeitssuche** und **Takterkennung** bietet. Datenauswertung ist aber auch im Zuge der computergestützten **Variantenermittlung** und Dokumentation mit der Editionssoftware *Edirom*¹¹⁴ möglich sowie der Anfertigung von aufführungspraktischem Material (Auszug von Stimmen, Anfertigung von Klavierauszügen, Chorpartituren usw.) und der Verknüpfung von Informationen aus unterschiedlichen Ebenen.

3.6 Theologie

Die Theologie ist eine Disziplin der Geisteswissenschaften, die sich mit vielfältigen Forschungsgegenständen und -fragen beschäftigt, interdisziplinär und international ausgerichtet ist und die zudem - entsprechend ihrer konfessionellen Ausrichtung und theologischen inhaltlichen Breite - aus einer Vielzahl von Teildisziplinen besteht, die mit Hilfe von unterschiedlichsten Methoden und Verfahren verschiedenste Erkenntnisinteressen und Forschungsfragen analysiert. So lassen sich beispielsweise Verfahren der Philologien, der Geschichtswissenschaft und auch der Kunstgeschichte, aber auch aus den Sozialwissenschaften (bei der Praktischen Theologie) finden, was die hohe Anschlussfähigkeit und hinsichtlich der Verfahren die sehr große Interdisziplinarität dieser geisteswissenschaftlichen Disziplin verdeutlicht. Aus diesem Grund soll im Folgenden vor allem auf die in den exegetischen Teildisziplinen vorzufindenden DH-Verfahren eingegangen werden. DH-Verfahren aus der Kirchengeschichte und den systematischen Teildisziplinen werden, da sie bereits in anderen Fachkontexten benannt wurden, nicht explizit ausgeführt.

Mit dem Index Thomisticus¹¹⁵, den Roberto Busa¹¹⁶ schon in den 1940er Jahren begonnen hat, und der als das erste Digital Humanities-Projekt überhaupt angesehen werden kann, ist in der Theologie eine der Wurzeln der Digital Humanities zu finden. Es existiert eine Reihe weiterer aktueller und vorbildlicher Projekte der Quellenerschließung und der Textkritik, wie z.B. der *Codex Sinaiticus*¹¹⁷ und das von der Universität Münster stammenden Projekt Digital-Nestle-

¹¹⁴ <http://www.edirom.de/>

¹¹⁵ <http://www.corpusthomisticum.org/>

¹¹⁶ http://de.wikipedia.org/wiki/Roberto_Busa

¹¹⁷ <http://www.codexsinaiticus.org/de/>

Aland¹¹⁸, eine digitalen Edition des Greek New Testament. Die Theologie hat vielfältige DH-Verfahren von anderen Fachdisziplinen in den letzten Jahren adaptiert, vor allem von der Computerlinguistik. Seit 20 Jahren gibt es die sogenannte Bibelsoftware, Softwarepakete – meist als CD-Editionen – die einfach zu nutzen und weit verbreitet sind. Diese Programme entsprechen aber aufgrund ihrer Abgeschlossenheit und vor allem aufgrund der Tatsache, dass diese CD-Editionen vielmehr für die pastorale Arbeit – also als Vorbereitungsmaterial für Predigten und für die Seelsorge – entwickelt wurden, nicht den Bedürfnissen und Erfordernissen von Wissenschaftlern, die eine digital gestützte Theologie betreiben wollen. Theologische Basistexte werden zwischenzeitlich durch webbasierende Angebote Wissenschaftlern zugänglich gemacht, so dass sie über “persistente kanonische Adressen” angesprochen werden können.¹¹⁹

Da die primären Datenquellen im Fach Theologie Texte sind, ist der Schwerpunkt im Bereich der Datenerfassung in erster Linie die Digitalisierung von Texten. Dieser umfasst nicht nur Scans und OCR von gedruckten Texten, sondern auch die **Transkription von Manuskripten**, die Beschreibung der Texte mithilfe von Metadaten und Normdaten, und die **Text-Bild-Verknüpfung** mit entsprechenden Faksimiles. Beispiele für Projekte, die digitale Texterfassung betreiben, sind: *The Cairo Genizah Manuscripts*¹²⁰ (Cambridge University), *Verbum: the Old Latin translation of the Gospel of John*¹²¹ (Universität Münster und Birmingham) sowie der vom Institut für neutestamentlichen Textforschung der Universität Münster erstellte *New Testament Virtual Manuscript Room*¹²².

Digital verfügbare Texte werden mittels **Kollation** oder **Stemmatologie** aufbereitet, also beispielsweise gemäß ihrer genetischen Relationen geordnet (z.B. Vorfahre-Nachkomme, Geschwister). Hierzu werden Softwareprodukte wie CollateX genutzt. Das Projekt *The Greek Bible in Byzantine Judaism* (GBBJ)¹²³ (King's College London und University of Cambridge) nutzt diese Technologie.

Im Zuge des **Encodings** werden Texte mit Zusatzinformationen bzw. Metadaten zum Dokument (Verfasser, Datum der Erstellung, Originalsprache etc.), z.T. auf Basis von Normdaten, angereichert oder mit Part-of-Speech-Tags für Wörter resp. mit Verweisen auf Informationen außerhalb des Dokuments – zumeist mit Hilfe von XML-Schemata bzw. TEI und

¹¹⁸ <http://nestlealand.uni-muenster.de/>

¹¹⁹ Z.B.: <http://www.bibleserver.com/text/VUL/Psalmen1,2>

¹²⁰ http://www.arts-humanities.net/projects/cairo_genizah_manuscripts_taylor_schechter_old_series_mosseri_collection

¹²¹ http://www.arts-humanities.net/projects/verbum_old_latin_translation_gospel_john

¹²² <http://intf.uni-muenster.de/vmr/NTVMR/IndexNTVMR.php>

¹²³ http://www.arts-humanities.net/projects/greek_bible_byzantine_judaism_gbbj

Software wie z.B. dem XML-Editor Oxygen. Solche sehr verbreiteten Techniken werden in einigen bereits genannten Projekten sowie den Projekten *The Co-edited Commentary on Augustine, City of God*¹²⁴ (University of Bristol) und *Controversia et Confessio* (Akademie der Wissenschaften und der Literatur, Mainz und IEG)¹²⁵ angewendet.

Namen von Orten, Personen und Organisationen innerhalb von Dokumenten werden durch Tools zur **Named-Entity-Extraktion** automatisch erkannt und mit Markup versehen. Darüber hinaus lassen sich Relationen zwischen Texten, Personen und Objekten durch **Geographic Information Software (GIS)** wie zum Beispiel *ArcGIS* visualisieren und analysieren. Als Beispiel für ein GIS nutzendes Projekt kann *The Digital Atlas of Roman and Medieval Civilization*¹²⁶ (DARMC; Harvard University) dienen, das geographische Daten der antiken und mittelalterlichen Welt verfügbar macht und analysiert.

Im Bereich digitaler Datenauswertung bieten **corpuslinguistische Verfahren** eine Vielzahl an weiteren Möglichkeiten, zum Beispiel die Extraktion linguistischer Informationen. Ein Beispielprojekt im Hinblick auf die Nutzung corpuslinguistischer Verfahren ist das bereits genannte Projekt *The Greek Bible in Byzantine Judaism* (GBBJ). Auch **Text Mining** ist ein oft genutztes digitales Verfahren zur Informationsextraktion, auch aus weniger wohldefinierten Textsammlungen – z.B. mit griechischen Texten aus dem 1. bis 4. Jahrhundert n. Chr. **Stylometrie** und Techniken zur **Zuordnung von Autorschaft** können zur Identifikation von Urhebern umstrittener Dokumenten angewendet werden. Text Mining sowie stylometrische Techniken werden zum Beispiel im Projekt *The Typology of Anonymous and Pseudepigraphic Jewish Literature in Antiquity, c. 200 BCE to c. 700 CE*¹²⁷ an der Durham University und der University of Manchester genutzt.

Die Entwicklung in der Theologie hin zu digitalen Verfahren ist vorhanden und gewinnt an Bedeutung, was z.B. daran zu sehen ist, dass es seit Kurzem im europäischen Teil der Society of Biblical Literature eine digitale Arbeitsgruppe gibt. In der Theologie sind schnelle Fortschritte möglich, weil ein sehr großer Teil der Primärtexte bereits in digitaler Form vorliegt.

3.7 Philosophie

Die Philosophie als akademisches Fach bewegt sich aktuell in der Nutzung von *DH-Verfahren* in eher grundlegenden Bereichen – elaboriertere Techniken werden durch ausserdisziplinäre

¹²⁴ http://www.arts-humanities.net/projects/co_edited_commentary_augustine_city_god

¹²⁵ <http://controversia-et-confessio.adwmainz.de/index.php?id=600>

¹²⁶ http://www.arts-humanities.net/projects/digital_atlas_roman_medieval_civilization_darMC

¹²⁷ http://www.arts-humanities.net/projects/typology_anonymous_pseudepigraphic_jewish_literature_antiquity_c_200_bce_c_700_ce

Do-it-yourself-Projekte mit aktuell eher wenig umgrenzten Forschungsfragen stärker illustrativ genutzt, bspw. in dem Projekt *Graphing the History of Philosophy*.¹²⁸ Die Identifikation aktueller, übergreifender *Hot Spots* der Nutzung elaborierterer *DH-Verfahren* im Fach ist insofern – abseits von bereits im Kontext der anderen Fächer erwähnten grundlegenden Verfahren wie Digitalisierung, Archivierung, allgemeineren Formen kollaborativen Arbeitens etc. – erschwert. Im philosophischen **Editionsbereich** existieren aktuell mehrere laufende Hybrid-Projekte (Jaspers, Kant, Leibniz, Lull, Nietzsche, Schopenhauer, Spinoza, Wittgenstein), von denen zu wünschen ist, dass Datenquellen in die Suche der virtuellen Fachbibliothek *Sophikon* integriert werden (vgl. zu philologischen *DH-Verfahren* allgemeiner auch 3.1). Projekte wie das *Digital Averroes Research Environment* des Thomas-Instituts, Köln – eine Forschungsumgebung (Edition, Inkunabeln, Bilder, Bibliographie) zum mittelalterlichen Philosophen Abū I-Walīd Ibn Rušd – sind in ihrer Komplexität eine Ausnahme. Wo in der Philosophie forschend mit Digitalisaten gearbeitet wird, werden Forschungstexte in **individuellen lokalen Arbeitsumgebungen mit jeweils verschiedenen spezifischen digitalen Komponenten** leichter handhabbar gestaltet. In Folge – insbesondere zugunsten der Relationierung digitaler Texte untereinander zugunsten jeweiliger Forschungsfragen – werden sie hierbei händisch integriert in eine hauptsächlich lokale Assemblage von Textverarbeitung und Bibliographielösung (Endnote, Citavi etc., oft offline-/standalone), teils auch einer Instanz des privat entstandenen elektronischen *Zettelkastens nach Luhmann* und/oder einer der vielfältigen, disziplinär passenden open-source *Mindmap*-Tools u.a. In einigen Fällen sind individuell zusammengefügte digitale Szenarien die Basis für Forschungs- oder Lehrsituationen, die experimenteller arbeiten, z.B. mit digitaler, teils kollaborativer Lektüre und digitaler Annotation bei möglichst anhaltender, unabgelenkter Konzentration der Rezipienten. Solche Forschungen finden an der TU Darmstadt im Kontext *Audioreader*¹²⁹ und *Leselabor*¹³⁰ statt. Am Rande philosophischer Forschung finden ähnliche Experimente mit Videos statt, so beim Projekt *lignes de temps*¹³¹ des Pariser IRI / Centre Pompidou. Daneben sind – stärker als in der analytischen Philosophie in der eher hermeneutisch geprägten Philosophie – manche andernorts einschlägige Verfahrensschritte quantitativer Kalkulation und Auswertung von Daten und deren Verortung im Forschungsprozess kaum selbstverständlich – zumindest potentiell absehbare Ausnahmen bilden aktuell editionsphilologische Kontexte, Begriffs- oder Metapherngeschichte. Wo hierbei in den individuellen elektronischen Arbeitsumgebungen philosophischen Forschung

¹²⁸ <http://drunks-and-lampposts.com/2012/06/13/graphing-the-history-of-philosophy/>

¹²⁹ <http://www.audioreader.org/Web-Site/willkommen..html>

¹³⁰ <http://www.leselabor.de/>

¹³¹ <http://www.iri.centrepompidou.fr/outils/lignes-de-temps/>

selbst auf digitale Quellen und Wörterbücher¹³² zurückgegriffen wird und weitgehend noch keine belastbaren Online-Lösungen existieren, wird auf digitale Lösungen zurückgegriffen wie das *historische Wörterbuch der Philosophie*, Datensätze liegen hierbei allerdings zumeist eher offline vor, als CD-ROM.¹³³

Relevante philosophische Forschung findet aktuell eher selten in speziellen VREs statt. In der philosophischen Forschung selbst geschieht – neben eher diversifiziert entstehendem, jeweils spezifischem „user-content“ – eine Anreicherung digitaler Forschungsdaten mit maschinenlesbaren Metadaten bislang weniger in digitalen Forschungsprojekten als bei individuellen Forschern oder in den Hybrid-Editionskontexten sowie philologischen Projekten. Für die Forschung relevant werden daneben aktuell z.B. pragmatische individuelle Experimente, beliebigen Nutzern die Möglichkeit zu geben, an der Lieferung von Metadaten – z.B. bibliographischen Verknüpfungen von Primärtexten – digital zu partizipieren, insbesondere rund um die *Spinoza-Bibliographie*¹³⁴ der Spinoza-Gesellschaft in Isernhagen.

3.8 Jüdische Studien¹³⁵

Als vielfältiges und interdisziplinäres international vernetztes Fach mit verteilten Forschungsstandorten, das die Judaistik im engeren Sinne, die Jüdischen Studien und die deutsch-jüdische Geschichte umfasst, aber auch auf angrenzende Disziplinen wie Hebraistik, Jiddistik, Religions- und Literaturwissenschaft sich erstreckt, spielt hier eine Fülle von Verfahren der DH aus den Philologien (s.o. 3.1) wie der Geschichtswissenschaft (s.o. 3.2) eine Rolle.

So etwa die **wissenschaftliche Textedition**, deren Ausgangspunkt meist die manuelle Texterfassung ist, zunehmend auch die aufwändig nachgearbeitete Texterfassung von in gebrochener Schrift gesetzten Büchern mittels OCR-Verfahren (Gesammelte Schriften Franz Oppenheimers, 1995ff., aktuell Deutsch-Jüdische Publizistik¹³⁶). Die so entstehende *Edition der Universal-Kirchenzeitung* berücksichtigt bereits bei der Datenaufbereitung die Interoperabilität mit Services wie PND und Georeferenzierung durch Referenzierung auf **Normdaten**.

¹³² <http://www.philosophie-woerterbuch.de/>; <http://woerterbuchnetz.de/DWB/>

¹³³ http://www.hwph.ch/index_schwabe.html

¹³⁴ http://www.spinoza-bibliografie.de/index.php?action=show_page_home

¹³⁵ Das folgende Kapitel zu „Jüdische Studien“ wurde maßgeblich vom Salomon Ludwig Steinheim-Institut für deutsch-jüdische Geschichte erstellt.

¹³⁶ <http://www.deutsch-juedische-publizistik.de>

Erwähnenswert insbesondere auch die synoptischen Editionen rabbinisch-jüdischer Traditionsliteratur.¹³⁷ Auf die Editionen der Hekhalot-Literatur folgten großangelegte Editionsprojekte zum Talmud Jeruschalmi und Talmud bavli. Computerunterstützter **Textvergleich, Kollation** und **synoptische Aufbearbeitung** wurden jeweils mit dem TUSTEP Modul "Vergleiche" erstellt. Parallel zu den synoptischen Editionen konnten aus den Quelldateien der synoptischen Editionen computergenerierte Konkordanz¹³⁸ generiert werden.

Hypertext als dem Quelltext – einem Traktat des babylonischen Talmuds – angemessene Präsentationsform sowie Eigenschaften einer **VRE** bot die Forschungsplattform *Comprehensive Study Environment for a Talmudic tractate (CESE)*. In der rabbinischen Auslegungstradition sind über Jahrhunderte zahlreiche intertextuelle "Schichten" zwischen Quellen und Kommentaren entstanden. Die Darstellung und Analyse dieser Beziehungen gelingt mittels DH-Verfahren besser und anschaulicher als in herkömmlichen Print-Editionen. Auch die Beziehungen zwischen den einzelnen rabbinischen "Schulen" oder mittelalterlichen Auslegungszentren könn(t)en mit der Einbeziehung von Geolocations in ihrem räumlichen und zeitlichen Zusammenhang leichter erforscht werden.

Am URN-Verfahren der Deutschen Nationalbibliothek nehmen Projekte wie die *Freimann-Sammlung*, die *Deutsch-Jüdische Publizistik* oder auch die Online-Ausgabe der Zeitschrift *Kalonymos* teil, so dass hier die Basiskatalogisierung, die Langfristverfügbarkeit und Langzeitarchivierung gesichert scheint. Die Beschränkung auf das hierfür bisher nur anwendbare PDF/A-Format ist jedoch für viele andere digitale Editionen nicht geeignet.

Der "Spatial Turn" nicht nur in den Jüdischen Studien geht einher mit dem zunehmenden Interesse an der digitalen Verarbeitung von Geolocations. Die Kartierung der Jüdischen Präsenz im Byzantinischen Reich mittels **GIS** unter Berücksichtigung sämtlicher Quellen hat sich das Cambridger Projekt *Mapping the Jewish Communities of the Byzantine Empire*¹³⁹ vorgenommen.

Das soeben bewilligte eHumanities-Projekt *GeoBib – Frühe deutsch- bzw. polnischsprachige Holocaust- und Lagerliteratur (1933-1949)* strebt eine "annotierte und georeferenzierte Online-Bibliografie zur Erforschung von Erinnerungsnarrativen" an. Ein „virtueller Atlas“ soll Gettos und

¹³⁷ Die Geschichte von den zehn Märtyrern: synoptische Edition mit Übersetzung und Einleitung, (Texte und Studien zum antiken Judentum; 10), hrsg. v. Reeg, Gottfried, Tübingen, 1985: In der Einleitung wird ausführlich auf die eingesetzten DH-Methoden und Verfahren eingegangen.

¹³⁸ Konkordanz zur Hekhalot-Literatur in Zusammenarbeit mit Gottfried Reeg u. unter Mitwirkung von Klaus Herrmann hrsg. v. Peter Schäfer, Tübingen, 2 Bände 1986, 1988.

¹³⁹ <http://www.mjcb.eu>

Lager nachweisen, Georeferenzierung die dort publizierten Texte mit weiteren auf Orte und Regionen bezogene Informationen verknüpfen, wie zum Beispiel mit Rezensionen, Sekundärliteratur und Bildquellen. Auch die *Central Database of Shoah Victims' Names* von Yad Vashem unterstützt seit neuestem Geolocations (googlemaps) und auch Ansätze kollaborativer Zusammenarbeit.

Ebenfalls in der Förderlinie eHumanities des BMBF widmet sich der interdisziplinäre Forschungsverbund *Relationen im Raum*¹⁴⁰, der Analyse und Visualisierung räumlicher Relationen zwischen Grabmalen jüdischer Friedhöfe aus neun Jahrhunderten und der Entwicklung eines "Topographie-Visualizer" zur **Visualisierung** topographischer Klein(st)strukturen.

Aufgrund der gezielten Zerstörung des jüdischen Schrifttums insbesondere durch den Nationalsozialismus spielt das Konzept der **virtuellen Bibliothek** in den Jüdischen Studien eine besondere Rolle. Hier geht es darum, Rara, zerstörte und verstreute Sammlungen zu rekonstruieren und online zugänglich zu machen und verschollenes Schriftgut wenigstens (online-) bibliografisch nachzuweisen. Nennenswerte Bestände sind weitaus früher digitalisiert worden, als dann später Standardisierungsbemühungen im Bereich Digitalisierung (s.o. 2.2) einsetzen: etwa das umfangreiche Zeitschriftenarchiv *compactmemory*, die Frankfurter *Sammlung jiddischer Drucke* oder etwa die *Exilpresse* der Deutschen Nationalbibliothek (die wegen potenzieller Urheberrechtsprobleme jüngst offline ging). In methodisch-technischer Hinsicht ist gerade die hier entstandene heterogene Vielfalt eine besondere Herausforderung für die Digital Humanities. So kann sich eine Judaica-Suchmaschine nicht auf eine ideale, homogenen Infrastruktur mit ihren Standard-Schnittstellen stützen, könnte und sollte aber als Prototyp zur Quellenvernetzung gerade dennoch entwickelt werden.¹⁴¹ Aktuelle Kooperationen wie die der *Judaica Europeana* eröffnen die Perspektive, **fachspezifische Recherchemöglichkeiten** auch mittels der *Europeana OpenSearch API* zu entwickeln.

Zumindest projektintern werden bei eher klassisch ausgerichteten Forschungen, die eine Printpublikation anstreben, Datenbanken zur Organisation und Analyse des Quellenmaterials benutzt.¹⁴² Mit seinem programmatischen Digital-Humanities-Bezug sucht das Projekt *The Haskala Republic of Letters Research Project* das Potenzial von **Semantic-Web-Technologien** für die Jüdischen Studien fruchtbar zu machen.¹⁴³ Hier wird die SQL-Datenbank eines

¹⁴⁰ <https://dev2.dariah.eu/wiki/display/RIRPUB/RiR>

¹⁴¹ http://steinheim-institut.de/wiki/index.php/Vieles_finden

¹⁴² So etwa das "Archiv jüdischer Kulturen" (Sächsische Akademie der Wissenschaften zu Leipzig), <http://www.saw-leipzig.de/forschung/projekte/europaeische-traditionen-enzyklopaedie-juedischer-kulturen>

¹⁴³ Testseiten: http://haskala.referata.com/wiki/Main_Page

Forschungsprojekts zur jüdischen Aufklärung¹⁴⁴ transformiert in ein *Semantic Media Wiki*¹⁴⁵ als "core tool" für das Vorhaben.¹⁴⁶ Ziel ist eine **kollaborative Forschungsplattform**, die Eigenschaften wie RDF, Linked Data, OAI-ORE sowie Ontologien und Annotationen unterstützt. Mit der "Codicological Data-Base of the Hebrew Palaeography Project"¹⁴⁷ hat das an der Jewish National and University Library (JNUL) angesiedelte *Institute for microfilmed Hebrew manuscripts* (IMHM)¹⁴⁸ in Jerusalem ein "tool for localizing and dating Hebrew Medieval Manuscripts" im Rahmen des 1968 begonnenen Hebrew Paleography Projects entwickelt, das auf Grundlage datierter Handschriften ein **automatisches Analyseverfahren zur Zuordnung von Schreibern, einer Datierung und Lokalisierung** undatierter hebräischer Handschriften bereitstellt.

Seit zehn Jahren vereint die epigraphische Datenbank jüdischer Grabinschriften epidat¹⁴⁹ **born digital** Editionen und **Retrodigitalisate** unterschiedlicher Provenienz in einem Inschriftenportal. Den epidat-Editionen stehen GeoBrowser, Indices und eine Ausgabe im kollaborativ entwickelten, auf TEI XML basierten **encoding** für epigraphische Daten zur Verfügung.¹⁵⁰ Intern wird **Entity Recognition** – Namen, Zitate, Orte, Daten, Idiomatik, Abkürzungen – für die strukturierte Inschriftenaufnahme angewandt.

http://atm.kcl.ac.uk/sites/atm.kcl.ac.uk/files/event/2011/09/semanticmediawiki_haskala.pdf

<http://atm.kcl.ac.uk/event/2011/09/26/judaica-europeana-haskala-jewish-enlightenment-database>

¹⁴⁴ Haskala: Building a modern Jewish Republic of Letters in the 18th and 19th Century.

¹⁴⁵ SMW ist die Open-Source-Extension der MediaWiki-Software.

¹⁴⁶ Vgl. http://www.judaica-europeana.eu/docs/ID_2-2_SMW_Haskala_Project_Draft.pdf.

¹⁴⁷ Malachi Beit-Arie: Codicological Data-Base of the Hebrew Palaeography Project: tool for localizing and dating Hebrew Medieval Manuscripts, <http://www.h-net.org/~judaic/oldlibrary/CODES.MBA.txt>

¹⁴⁸ <http://jnul.huji.ac.il/imhm/>

¹⁴⁹ <http://steinheim-institut.de/cgi-bin/epidat>

¹⁵⁰ <http://epidoc.sourceforge.net/>

4. Fazit: Kern-DH-Verfahren als disziplinenübergreifende Perspektive für die Entwicklung der DARIAH-DE-Infrastruktur

Die überblicksartige Zusammenstellung der aktuellen *Hot Spots* in den verschiedenen geistes- und kulturwissenschaftlichen Disziplinen zeigt, dass disziplinübergreifend Überschneidungen bei der Verwendung bestimmter *DH-Verfahren* festgestellt werden können. Um weitere Anschlussmöglichkeiten und in der DARIAH-DE-Infrastruktur umsetzbare Bedarfe übergreifend zu identifizieren, werden im Folgenden disziplinäre Gemeinsamkeiten und Unterschiede in der Verwendung von *DH-Verfahren* gebündelt.

Editions- bzw. Erschließungsverfahren werden in allen untersuchten Fachdisziplinen verstärkt computergestützt verwendet. Dies gilt unabhängig davon, ob es sich um eher textwissenschaftliche Fächer handelt, in denen beispielsweise editionsphilologisch oder historisch-kritisch geforscht wird, oder um eher bild- bzw. objektorientierte Disziplinen wie Kunstgeschichte oder Archäologie, oder auch um die Musikwissenschaft, in der an Werkeditionen, Noteneditoren und dem MEI-Standard gearbeitet wird. Hier werden jeweils relevante Daten, die in den spezifischen Forschungsgegenständen enthalten sind bzw. von ihnen repräsentiert werden, erfasst, transkribiert und aufbereitet, um einerseits verschiedene Formen der Dokumentation bzw. Bestandserfassung mit Informationen anzureichern (z.B. durch Annotation und Kontextualisierung), andererseits analytischen Zugriff auf die Forschungsgegenstände zu ermöglichen bzw. zu erweitern.

Besonders stark ausgeprägt sind Überschneidungen in den schwerpunktmäßig textwissenschaftlichen Disziplinen, in denen größtenteils ähnliche oder gar identische Editionsverfahren und auch -methoden angewendet werden. Für bild- bzw. objektorientierte Disziplinen angewandte typische Verfahren gewinnen aber auch in philologischen und historischen Fächern in einem erkennbaren Maße an Bedeutung (z.B. für die Untersuchung von Textträgern wie Handschriften). Die Verknüpfung von Daten und Forschungsgegenständen mit ihren Kontexten wird aktuell vorrangig innerdisziplinär umgesetzt (z.B. Verknüpfung von digitalisierten Texten mit digitalen Wörterbüchern, Erschließung von Relationen in Briefnetzwerken usw). Durch Metadaten und gemeinsame Standards werden interdisziplinäre Vernetzungen zunehmend erst ermöglicht. Mehrwerte werden im Zuge von Editions- bzw. Erschließungsverfahren vor allem durch das Verfügbarmachen für den analytischen Zugriff –

auch interdisziplinär – sowie in Form von Kontextualisierung und Verknüpfung erzeugt; dies sind Verfahren, deren digitale Umsetzung und Verfolgung in größerem Maßstab erst durch eine entsprechende Infrastruktur gewährleistet werden können. Möglichkeiten zur Langzeitarchivierung und dauerhaften Präsentation sind hierbei Voraussetzung für eine nachhaltige Erstellung digitaler Editionen bzw. Sammlungen von digital erschlossenen bzw. zu erschließenden Forschungsgegenständen, eine Notwendigkeit, die bislang in der deutschen Forschungslandschaft gerade für Forschungsdaten noch nicht genügend berücksichtigt wird.

Computergestützte Analyseverfahren sind insgesamt stärker fachspezifisch geprägt. Sie werden durch die jeweiligen Untersuchungsgegenstände, die daran gerichteten Forschungsfragen und deren Verständnis sowie deren vorherige, jeweils mehr oder weniger fachspezifisch geprägte Erschließung bestimmt. Aber auch auf dieser Ebene gibt es einige Disziplinen übergreifende Entwicklungen. So werden in mehreren Disziplinen gerade die *DH-Verfahren* genutzt, die die Anwendung quantitativer Analysen sowie qualitativ-semantischer Verfahren auf große (Text-) Datenmengen ermöglichen bzw. unterstützen. Als ein in unterschiedlichen Kontexten etabliertes und mit verschiedenen Techniken umgesetztes *DH-Verfahren* kann die **Mustererkennung** betrachtet werden. Muster können dabei in einem weiteren Sinn nicht nur als visuelle, sondern auch als sprachliche Muster bzw. Strukturen verstanden werden, wobei sich das Erkennen von Mustern in Bildern hinsichtlich der Prozessierung etc. allerdings stark von dem in Texten unterscheidet. Insofern werden Mustererkennungsverfahren in verschiedenen Varianten in den einzelnen Disziplinen genutzt, z.B. in Form von Text Mining-Verfahren (wie Stylometrie und Topic Modeling) in den Philologien bzw. in den textwissenschaftlichen Bereichen der anderen Fächer, in Form von Netzwerkanalysen, Kartierungen oder auch im Rahmen von Visualisierungen, die Muster sichtbar und somit analysierbar und vergleichbar machen. Muster können ein unterschiedliches Abstraktionsniveau bzw. einen unterschiedlichen Grad an Komplexität aufweisen, der das konkrete Vorgehen innerhalb des Verfahrens bestimmt.

Es ist naheliegend, hier ein großes Potenzial für einen interdisziplinären Austausch anzunehmen. Parallelen im Bereich von Analyseverfahren zeigen einen Bedarf an interdisziplinärem Austausch (z.B. hinsichtlich Mustererkennung in Texten, graphischen Elementen, Zeichen usw. bezogen auf unterschiedliche Forschungskontexte) mittels einer digitalen Forschungsinfrastruktur wie DARIAH-DE. Dies gilt nicht nur hinsichtlich der Struktur und eines optionalen Einsatzes von solchen *DH-Verfahren* bzw. entsprechender Tools, sondern auch für den interdisziplinären Austausch von Forschungsdaten.

Im Bereich der **Visualisierung** lassen sich zwei Stränge ausmachen: Zum einen das Visualisieren von Zusammenhängen etc. auf Grundlage von erfassten und aufbereiteten Daten, zum anderen die (Re-)Konstruktion von Objekten, Bauwerken etc. in Form von **3D-Modellen**. Beide Formen der Visualisierung werden einerseits als Darstellungsverfahren eingesetzt, andererseits aber auch als ein Verfahren zum Erkenntnisgewinn. Insofern handelt es sich bei der Visualisierung um ein Verfahren, das bei der Datenaufbereitung sowie der Dokumentation und Publikation eine immer größere Rolle spielt, aber auch um ein Analyseverfahren, das für die Datenauswertung genutzt wird, indem Relationen und Muster digital gestützt sichtbar gemacht werden, die ohne Computertechnik nicht zugänglich wären.

Hierbei spielen insbesondere Visualisierungstechniken eine Rolle, die es ermöglichen, große Datenmengen, die bislang durch Einzelforscher aufgrund der ungeheuren Datenmasse nicht wissenschaftlich untersucht werden konnten, zu analysieren. Diese neuartigen DH-Visualisierungsverfahren – zu denen beispielsweise auch der DARIAH-DE Geobrowser zählt – zeigen sehr deutlich, dass Forschungsdaten nicht nur interdisziplinär genutzt und nachgenutzt werden, sondern auch, dass bestimmte Verfahren der Digital Humanities transdisziplinär eingesetzt werden können. Ebenso von Bedeutung sind Forschungsinfrastrukturen im Zusammenhang mit Visualisierungen – sowohl bezüglich des Zugriffs auf visualisierbare Datenbestände als auch hinsichtlich der Publikation in dynamischer Form, die erst den Mehrwert von Visualisierungen zugänglich macht.

Die **Anwendung und Nutzung von Normdaten** ist ein *DH-Verfahren*, das sich aktuell in fast allen in diesem vorliegenden Report analysierten Fachdisziplinen findet, unabhängig davon, ob es sich eher um digitale Erfassungs- oder Digitalisierungsverfahren handelt oder um Forschungsgegenstände, die eher durch einen klassisch hermeneutischen Zugang charakterisiert sind. Die Verwendung von Normdaten (standardisierte Verweise auf Personen, Orte, Ereignisse, etc.) hat für *DH-Verfahren* deshalb eine besondere Bedeutung: Durch sie kann sichergestellt werden, dass interoperable Nachnutzungsszenarien für Forschungsdaten und -ergebnisse entstehen, um so einen disziplinübergreifenden interoperablen Zugriff zu ermöglichen. DARIAH-DE stellt mit dem Archiv-Editor des Personendatenrepositoriums (PDR), das derzeit von der BBAW entwickelt und erprobt wird, ein Tool zur Erstellung/Bearbeitung von Personendaten bereit, bei dem zukünftig auch Personennormdateien integriert werden sollen. Über entsprechende Schnittstellen ist hier beispielsweise an eine Verlinkung mit dem DARIAH-DE-GeoBrowser, einem weiteren DARIAH-DE-Demonstrator, denkbar. Dies wäre ein exzellentes Beispiel für eine interoperable Nachnutzung von Forschungsdaten

unterschiedlichster Provenienz und aus unterschiedlichster Forschungsperspektiven, die durch die Anreicherung mit Normdaten einen qualitativen, substanziellen Mehrwert erhalten und auf diese Weise maschinenlesbar und interoperabel werden können. Forschungsdaten können auf diese Weise mit unterschiedlichsten Informationsquellen, mit aktueller Forschungsliteratur und einer beliebigen Zahl von digitalen Datensammlungen und Forschungsdaten verknüpft werden, um auf diese Weise neue, bislang von einzelnen Disziplinen nicht oder nur kaum beachtete Quellen und Daten in eigene Forschungskontexte einzubeziehen. Daraus entstehen neue Forschungsdimensionen, die eben nur aufgrund der Verwendung von *DH-Verfahren* realisiert werden können. Verallgemeinert gesprochen bilden die Normdaten ein wichtiges Fundament für eine allgemeine und weiter ausgreifende **semantische Annotation und Erschließung**. Diese eröffnet neue Perspektiven der Vernetzung und Nachnutzung digitaler Ressourcen. Sie ist vor allem aber die Grundlage für neue systematische analytische Untersuchungen (s.o.).

Neue Möglichkeiten im Umfeld der *DH-Verfahren* sind zuletzt auch auf der Ebene von **Kollaborations-Szenarien** sowie neuen **Publikations- und Kommunikationskulturen** zu erwarten. Hierzu ist auch die kollaborative Produktion dynamisch vernetzter Texte bzw. Datenstrukturen (Hypertext/Hypermedia) zu zählen. Auch diese Verfahren können von vertrauenswürdigen und robusten digitalen Forschungsinfrastrukturen profitieren, wenn sie diese in bestimmten Fällen nicht sogar voraussetzen. Die Entwicklung und Etablierung virtueller Forschungsumgebungen ist in diesem Zusammenhang von besonderer Bedeutung. Hierzu ist es allerdings notwendig, dass sich in den einzelnen geisteswissenschaftlichen Fachdisziplinen eine grundlegende Akzeptanz dieser neuen digital gestützten Verfahren und digitalen Publikationsmöglichkeiten einstellt. Nur wenn digitale Publikationen zukünftig als vollwertige wissenschaftliche Leistungen und als forschungsrelevant in den einzelnen Geisteswissenschaften anerkannt werden, wird sich eine nachhaltige Veränderung in den Publikations- und Kommunikationskulturen etablieren. In diesem Kontext wäre es wünschenswert, dass DFG und BMBF bei diesem Prozess die Federführung übernehmen und zukünftig verstärkt auch rein digitale Publikationen finanziell fördern.

Als ein zentrales **Ergebnis des vorliegenden Reports** für die Weiterentwicklung der DARIAH-Infrastruktur ist folgendes festzuhalten: Den vorgestellten *DH-Verfahren* – unabhängig davon, ob sie als *Hot Spots* in mehreren Fachdisziplinen vorzufinden sind oder nicht – ist gemein, dass sie auf vergleichbare technische Anforderungen stoßen, die sich einerseits auf die langfristige

Verfügbarkeit von Forschungsdaten, andererseits auf die Möglichkeiten der Anbindung an vielfältige Werkzeuge der Weiterverarbeitung und Analyse beziehen.

So gilt für alle Disziplinen, dass digital vorliegende Forschungsdaten die Grundlage für die Behandlung von Forschungsfragen und damit auch für die vielfältigen in diesem Report vorgestellten und evaluierten Verfahren der Digital Humanities sind. Es muss sichergestellt sein, dass Forschungsdaten nicht nur der Nachnutzung durch Dritte zur Verfügung gestellt werden können – was bereits bei deren Erhebung und digitalen Weiterverarbeitung beachtet werden muss – sondern dass auch erprobte und einsatzfähige Konzepte der Bitstream Preservation und Langzeitarchivierung zur Verfügung stehen. Dabei meint Langzeitarchivierung im weiteren Sinne auch die Gewährleistung einer andauernden Präsentation und Zugänglichkeit von Projektdaten und Projektergebnissen. Beides basiert auf einem umfassenden Forschungsdatenmanagement, das die spezifischen Eigenheiten von Forschungsdaten aus geisteswissenschaftlichen Kontexten berücksichtigen muss. Den geisteswissenschaftlichen Forschern muss außerdem technisch ermöglicht werden (und dass muss konzeptionell vorbereitet und politisch/sozial im Bewusstsein der Forschung durchgesetzt werden), Forschungsdaten zu referenzieren – z.B. durch PID-Systeme – und diese auf Dauer verwenden zu können. Um diese Forschungsdaten auch interdisziplinär und aus verschiedenen Perspektiven nutzen zu können, müssen zudem elaborierte Suchmöglichkeiten vorhanden sein, denn nur dann können Forschungsdaten auch aufgefunden und aus interdisziplinärer Perspektive mit weiteren Metadaten und Normdaten angereichert werden. DARIAH-DE bietet mit seiner *Generischen Suche* in Verbindung mit *Collection Registry* und *Schema Registry* ein Bündel an Funktionalitäten und Basisinfrastrukturen, mit deren Hilfe höchst unterschiedliche Forschungsdaten- und Quellentypen, die für geisteswissenschaftliche Forschung unabdingbar sind, vom einzelnen Forscher genutzt und nachgenutzt werden können.

Kern der Infrastruktur sollte dabei sein, Datensätze (mittels Repositorien, Archiven, Registries) und jene vielfältigen Tools miteinander zu verknüpfen, die die im vorliegenden Report als zentral markierten *DH-Verfahren* im engeren Sinn unterstützen. Es müssen zu diesem Zweck zudem nachhaltige, auf Dauer angelegte Infrastrukturen für das Hosting von Daten sowie Schnittstellen von und zu Datenrepositorien existieren. Jenseits der Bedürfnisse nach einer generischen IT-Basisinfrastruktur ist zu erwarten, dass der Bedarf an spezifischen Tools bzw. der Anbindung solcher Tools zur Datenanalyse an bestehende Forschungsdatenbanken steigen wird – wodurch dann eine direkte, flexibel an die spezifischen Bedürfnisse angepasste Auswertung von Daten vorgenommen werden kann. Ein erster Schritt in diese Richtung wird in

DARIAH-DE derzeit mit der Anbindung des *TextGridRep* an die *Voyant Tools* unternommen, weitere Schritte sind wünschenswert und im Kontext der 'Demonstratoren' absehbar.

DH-Verfahren werden hinsichtlich ihrer Verbreitung und ihrem Nutzwert deutlich von einer nachhaltigen Basisforschungsinfrastruktur profitieren, wie sie gegenwärtig von DARIAH-DE geplant und auf europäischer Ebene realisiert werden soll. Sie gewährleistet, dass die mit Hilfe der beschriebenen *DH-Verfahren* – seien sie nun disziplinären oder generischen Charakters – erhobenen Forschungsdaten und jeweils generierten Forschungsergebnisse nicht nur digital vorliegen, sondern auch nachhaltig digital nachgenutzt, mit bestehenden oder neuen Tools untersucht, und unter neuen DH-Forschungsfragen analysiert werden können. Die tiefgreifenden Entwicklungen, die die Geistes- und Kulturwissenschaften im Moment und in absehbarer Zukunft im Zuge der digitalen Durchdringung durchlaufen, müssen auf diese Weise durch Forschungsinfrastrukturen bestmöglich unterstützt werden.