

**Humboldt-Universität zu Berlin**

**Institut für Bibliotheks- und Informationswissenschaft**

**ABSCHLUSSARBEIT**

**Virtuelle Forschungsumgebungen als Entwicklungsfeld für Bibliotheken am  
Beispiel des „Deutschen Textarchivs“**

**Philosophische Fakultät I**

**Julia Elisabeth Klein**

**Gutachter: 1. Prof. Dr. Stefan Gradmann  
2. Prof. Dr. Manfred Thaller**

**Datum der Einreichung: 18.05.2012**

# Inhaltsverzeichnis

1. Einleitung.....	4
1.1 Grids.....	10
1.2 Virtuelle Forschungsumgebungen.....	12
2. Digital Humanities – Geisteswissenschaft im digitalen Raum.....	16
2.1 Grundlagen wissenschaftlichen Arbeitens in den Geisteswissenschaften.....	19
2.2 Neue Arbeitsformen in den Geisteswissenschaften.....	24
3. Virtuelle Forschungsumgebungen als Entwicklungsfeld.....	30
3.1 Wissenschaftliche und institutionelle Anforderungen.....	31
3.2 Technische Anforderungen.....	35
3.3 Linked Data als Chance für interpretationsgetriebene Aussagesysteme.....	37
4. Das „Deutsche Textarchiv“ – vom digitalen Korpus zur Virtuellen Forschungsumgebung.....	42
4.1 Korpusbeschreibung.....	43
4.2 Forschungswerkzeuge.....	44
4.2.1 Sucheinstiege.....	44
4.2.2 Linguistische Suchmaschine.....	51
4.3 Geplante Weiterentwicklung.....	53
4.3.1 Aktives Archiv.....	53
4.3.2 Parametrisierbare Leseansicht.....	54
4.3.3 Webbasierte Korrekturumgebung.....	55
4.3.4 Erweiterungsansätze.....	57
5. Virtuelle Forschungsumgebungen als Entwicklungsfeld für Bibliotheken.....	59
5.1 Sammlungen, Daten, Wissen und Werkzeuge.....	59
5.2 Generische Virtuelle Forschungsumgebungen und Forschungspartnerschaft.....	62
5.3 Die Zukunft des Fachreferats.....	66
6. Bibliotheken und Virtuelle Forschungsumgebungen.....	67
Literaturverzeichnis.....	70

## Abbildungsverzeichnis

Abbildung 1: Zuordnung wissenschaftlicher Primitive zu Wissenschaftsparadigmen nach Palmer et al. (2009: 35).....	22
Abbildung 2: Zentrale Anforderungskomplexe an Virtuelle Forschungsumgebungen.....	31
Abbildung 3: Startseite des Deutschen Textarchivs.....	45
Abbildung 4: Liste der zugreifbaren Bücher des DTA.....	46
Abbildung 5: Zeitleiste.....	47
Abbildung 6: Ansicht des Einzeltitels „Der Geisterseher“ von Friedrich Schiller im DTA-Viewer.....	48
Abbildung 7: Einfache Suchanfrage nach Geisterseher mit Anzeige aller Fundstellen.....	49
Abbildung 8: Textauswahl mit individueller Kategorisierung und Annotation.....	50
Abbildung 9: Funktionen des DTA-Viewers.....	51
Abbildung 10: Zukünftige parametrisierbare Leseansicht.....	55
Abbildung 11: Webbasierte Korrekturumgebung.....	56
Abbildung 12: Generische Virtuelle Forschungsumgebung.....	64

## Tabellenverzeichnis

Tabelle 1: wissenschaftliche Primitive nach Unsworth (2000), Palmer et al. (2009) und Blanke/Hedges (2011) im Vergleich.....	23
Tabelle 2: Anforderungen an Virtuelle Forschungsumgebungen.....	37

## 1. Einleitung

Die empirische Linguistik kann als die naturwissenschaftlichste der klassischen Geisteswissenschaften betrachtet werden. Als eine der ersten dieser traditionellen Disziplinen versuchte sie, ihren Untersuchungsgegenstand Sprache mit den exakten Methoden der Naturwissenschaft zu erfassen. Diese Aufgeschlossenheit gegenüber neuen Technologien bewahrt sich die Linguistik bis ins heutige Computerzeitalter und beginnt auch, sich an die Möglichkeiten des Forschens im Cyberspace heranzuwagen. Verfahren zur quantitativen Dialektalitätsmessung oder zur diskursanalytischen Betrachtung innovativen Sprachgebrauchs kommen heute ohne computergestützte Analyseverfahren in Webanwendungen nicht mehr aus. So bot etwa die Volltextdatenbank COSMAS I bzw. COSMAS II des Instituts für Deutsche Sprache als eine der ersten durch ihre zugehörigen Applikationen die Möglichkeit, quantitative Analysen in großen volltextdigitalisierten Textkorpora vorzunehmen.<sup>1</sup> Vor allem die korpuslinguistische Forschung der empirischen Linguistik zeigt, dass sich durch moderne Computertechnik völlig neue, nie gekannte Perspektiven bei der Analyse riesiger Korpora eröffnen. Der Weg der Linguistik vom analogen Schreibtisch in den digitalen Raum steht dabei exemplarisch für den Weg einer geisteswissenschaftlichen Disziplin in den Cyberspace. Mit der rasanten Weiterentwicklung der Informations- und Kommunikationstechnologie haben sich im vergangenen Jahrzehnt auch die technologischen Grundlagen und Möglichkeiten des Forschens verändert. Durch die zunehmende Leistungsfähigkeit von Datenleitungen, die erhöhte Kapazität moderner Speichermedien bis weit in den Petabyte-Bereich hinein und die Entwicklung komplexer Infrastrukturnetze, die eine effiziente Ressourcennutzung auf verteilten Rechnersystemen erlauben, können heute immer größere Datenmengen mit immer rechenintensiveren Operationen, die an verschiedenen geographischen Orten gespeichert sind, zeitlich synchron bearbeitet werden. Diese Entwicklung geht auch an den Geisteswissenschaften und Informationsdienstleistern nicht spurlos vorbei. Immer größere Datenkorpora aus zunehmend anspruchsvolleren Quellen mit immer mehr Ansatzpunkten zu deren

<sup>1</sup> Aktuell wird der Umfang des COSMAS-Korpus mit 5,4 Mrd. Wortformen (tokens) beziffert (vgl. Institut für Deutsche Sprache, 2003; Institut für Deutsche Sprache, 2010a). Bei COSMAS II handelt es sich jedoch um eine Applikation, die derzeit nur als Softwareclient, Webservice oder Kommandozeileninterpreter, nicht aber als vollständige Arbeitsumgebung ähnlich einer integrierten Entwicklungsumgebung vorliegt (vgl. Institut für Deutsche Sprache, 2010b).

Erforschung scheinen von Forschern gewünscht, digitale Volltexte aus Frakturdrucken<sup>2</sup> und Volltextdatenbanken aus Erhebungen der qualitativen Sozialforschung sind typische Beispiele dafür. Diesen neuen Anforderungen der Wissenschaft an Datengrundlagen und Informationsdienstleistung müssen sich auch wissenschaftliche Bibliotheken stellen, wenn sie nicht in den Nimbus des zwar Altherwürdigen, aber Überkommenen abgleiten wollen. Viele Bibliotheken haben dies bereits für sich als Chance erkannt und vollziehen den Übergang von einer Institution, die sich mit dem Sammeln, Erschließen und Bereitstellen von Information begnügt, zum kompetenten Partner der Forschung, der gemeinsam mit Wissenschaftlern Datenbasen erarbeitet, zur Beforschung aufbereitet und die entstehenden Forschungsdaten verwaltet.<sup>3</sup> Doch das potentielle Entwicklungsfeld für Bibliotheken erstreckt sich weit über die Digitalisierung von Texten und das Management von Primärforschungsdaten hinaus: mit großen Mengen von Forschungsdaten sind zu deren Bearbeitung heute auch die Virtuellen Forschungsumgebungen verbunden, die infrastrukturell geplant, aufgebaut, für den konkreten Forschungsbedarf angepasst, gepflegt und weiterentwickelt werden müssen.

Die Forderung nach leistungsfähigen Infrastrukturen, die ressourcenintensives Arbeiten erlauben, entstand zuerst in den Naturwissenschaften, vor allem in der Astro- und Hochenergiephysik, in der es galt, die immensen Datenmengen, die bei Versuchen mit Radioteleskopen und Teilchenbeschleunigern entstanden, auszuwerten. Die Forderung nach einer technischen Infrastruktur, die Wissenschaftlern einen globalen Zugang zu diesen großen Datenmengen und den Rechenressourcen zu ihrer Analyse bietet, brachte einen neuen Komplex aus Methoden, Konzepten und Technologien, in dem sich wissenschaftliches Arbeiten besonders in den Naturwissenschaften seither bewegt (vgl. Aschenbrenner et al., 2007, S. 11f). Dies vereint eine technologische mit einer sozialen Entwicklung des

---

2 Der Entwicklung der Volltext-Digitalisierung von Fraktur-Schriften widmet sich seit einigen Jahren das europäische IMPACT-Projekt (Impact, 2012), an dem Forschungseinrichtungen, Bibliotheken und Privatunternehmen beteiligt sind.

3 So kooperiert die Staats- und Universitätsbibliothek Bremen beispielsweise mit dem Fachbereich Sozialwissenschaften im Rahmen des „Grenzboten“-Projekts (Staats- und Universitätsbibliothek Bremen, 2011) bei der Erstellung eines volltextdigitalisierten Korpus der gleichnamigen Zeitschrift aus dem 19. Jahrhundert.

Forschungsprozesses. Auf der technologischen Ebene entstanden zunächst die sogenannten *Grids*.

Im Grid zu arbeiten bedeutet einerseits das global verteilte Arbeiten verschiedener Akteure und andererseits deren soziale Rollen im Forschungsprozess zu definieren und in Abhängigkeit von diesen sozialen Rollen technische Zugriffsrechte zu verwalten. Es zeigt sich also, dass sich mit dem Entstehen der Grids und ihrer neuen Möglichkeiten nicht nur neue Methoden, sondern auch neue Verhaltens- und Kommunikationsweisen wissenschaftlichen Arbeitens entwickelten. Dadurch haben Grids über ihre technologische Funktion des Bearbeitens von Datenmengen hinaus mit der Etablierung virtueller Organisationen auch eine soziale Ebene entwickelt. Diese Veränderung zeigt sich vor allem in der Evolution einzelner Arbeitsgruppen zu vernetzten virtuellen Organisationen. Dabei handelt es sich um verteilte Wissenschaftlergruppen, die im Extremfall aus einzelnen Personen oder Arbeitsgruppen unterschiedlichen Ländern oder Kontinente bestehen, und durch Computertechnologien an den dispersen Korpora arbeiten (vgl. Aschenbrenner et al., 2007, S. 12). Das Arbeiten in solchen virtuellen Strukturen erfordert eine technische Unterstützung, die die gemeinsame Bearbeitung der Daten, ihres Austauschs, die Kommunikation darüber und günstigstenfalls auch die Publikation der Ergebnisse auf komfortable Weise ermöglicht. Vor dem Hintergrund dieser Anforderungen, die ursprünglich auf die naturwissenschaftliche Forschung zurückgehen, entwickelte sich die Idee, diese Dienste in Form einer integrierten Entwicklungsumgebung (IDE)<sup>4</sup> zu realisieren, die den Zugriff auf die benötigten Forschungsdaten mit Software-Tools zu deren Beforschung, Verwaltung und Speicherung sowie Werkzeugen zur vernetzten Kommunikation und späteren Publikation ermöglichen. Die so entstandenen Virtuellen Forschungsumgebungen sollen diese Services bieten und haben sich mittlerweile auch weit über die Grenzen der naturwissenschaftlichen Forschung hinaus verbreitet. Daher halten sie auch zunehmend Einzug in die methodisch und kommunikativ stark vom Herkunftsparadigma der virtuellen Forschungsumgebungen abweichenden Geistes- und Sozialwissenschaften.

---

4 Die Abkürzung basiert auf dem Terminus *integrated development environment*.

Was für das naturwissenschaftliche Paradigma<sup>5</sup> bereits seit geraumer Zeit alltäglich ist, entwickelt sich seit einigen Jahren unter dem Schlagwort *eHumanities* zu einer Herausforderung für die Geisteswissenschaften: der Sprung in eine Welt, in der Forschung mit verteilten Ressourcen und Rollen global vernetzt an digitalen Schreibtischen in virtuellen Forschungsgruppen im Cyberspace stattfindet. Diese kulturelle Wende in der Forschungsarbeit erfordert zum Einen die Verknüpfung zwischen eigentlichem Forschungsprozess und informationstechnischer Infrastruktur, die das Arbeiten in einem digitalen Raum überhaupt erst ermöglicht. Zum Anderen bedeutet sie eine drastische Veränderung geisteswissenschaftlichen Arbeitens selbst, indem die Konsultation anderer Forscher bei der Bearbeitung spezifischer Themenbereiche deutlich vereinfacht und gleichzeitig kollaboratives Arbeiten als neue Forderung von außen an die Geisteswissenschaften herangetragen werden.<sup>6</sup>

Diese Weiterentwicklung wird auf internationaler Ebene gefördert, aber auch gefordert. In Großbritannien befasst sich das nationale Joint Information Systems Committee (JISC) mit dem Aufbau der Forschungsinfrastruktur, die für Forschung im virtuellen Raum notwendig ist. In Deutschland werden die Entwicklungen im Bereich der Forschung und Informationsdienstleistung von der Deutschen Forschungsgemeinschaft (DFG) gefördert. Die Vision der DFG bezieht sich darauf, dass jeder Forscher künftig von seinem Schreibtisch aus Zugang zu allen Informationen haben soll, die er für seinen Forschungsprozess benötigt, seien es Publikationen, Primärforschungsdaten oder Kommunikationsmittel. Dahinter steht die implizite Auffassung, Forschung werde verbessert und der Forschungsprozess dadurch vereinfacht, dass Primärforschungsdaten sowie Werkzeuge zu ihrer Analyse in einer einzigen digitalen Arbeitsumgebung verfügbar und bearbeitbar sind (vgl.

5 Die Begriffe *geisteswissenschaftliches* und *naturwissenschaftliches Paradigma* beziehen sich auf die grundsätzlich verschiedene Auseinandersetzung dieser Disziplinenkonglomerate mit ihren Gegenständen. Die Naturwissenschaften gehen grundsätzlich von der Existenz einer objektiven, empirisch nachprüfaren Realität aus, deren Messung logische Deduktion bzw. Induktion zulässt, um weitere Aussagen über die empirische Realität zu ermöglichen. Sie stellen in diesem Sinne *a-posteriori-Wissenschaften* dar. Dagegen basieren Aussagesysteme in den Geisteswissenschaften auf der interpretativen Auseinandersetzung mit einem Realitätsausschnitt vor dem Hintergrund verschiedenster Realitäts-, Wissens- und Weltkonzepte und können daher im hier referenzierten Kontext als *a-priori-Wissenschaften* gelten. Aufgrund dieses konzeptionellen Unterschieds in der Wahrnehmung von Welt und Wirklichkeit werden diese Konzepte im folgenden als geistes- bzw. naturwissenschaftliches Paradigma bezeichnet.

6 Diese Entwicklung wird in Kapitel 1.3.1 eingehender beleuchtet.

Sahle, 2008, S. 1). Dies soll durch verschiedene Förderprogramme zum Auf- und Ausbau wissenschaftlicher Informationsinfrastruktur ermöglicht werden. Zu diesem Zwecke entfielen im Jahr 2009 allein 63 Millionen Euro des damals 2,2 Milliarden umfassenden Gesamtbudgets der DFG auf die Förderung wissenschaftlicher Informationsinfrastruktur. Die Förderung des Aufbaus Virtueller Forschungsumgebungen und der dazugehörigen Primärforschungsdaten gehört dabei zu den zentralen Förderelementen (vgl. Deutsche Forschungsgemeinschaft, 2011b, Deutsche Forschungsgemeinschaft, 2009).<sup>7</sup> Designiertes Ziel der Virtuellen Forschungsumgebungen ist aus Sicht der DFG die Vereinfachung wissenschaftlicher Forschung, Kommunikation und des Publizierens (vgl. Lipp, 2009; Meyer, 2011, S. 39), die jedoch nur durch die Kooperation von Informations- und Forschungseinrichtungen geleistet werden können.<sup>8</sup> Darüber hinaus wird ein „weiterer wesentlicher Erfolgsfaktor [...] darin gesehen, dass sowohl bei Konzeption oder Aufbau neuer als auch der Weiterentwicklung vorhandener Strukturen gewährleistet ist, dass diese sich interoperabel in internationale Netzwerke einbinden lassen“ (Deutsche Forschungsgemeinschaft, 2010, S. 2). Im Rahmen dieses Kooperationsgedankens bietet die DFG Bibliotheken Fördermöglichkeiten für die technische Verbesserung von Informations-, Kommunikations- und Partizipationservices (im Sinne etwa von web 2.0-Diensten), um Informationsstrukturen auszubauen. Um den Nutzen langfristig zu garantieren, muss die Nachnutzbarkeit der Ergebnisse sichergestellt werden. Dies soll durch die Schaffung generischer, offener Systeme, die sich an Interoperabilität und offenen Standards orientieren und die ihre Grundlagen quelloffen zur Verfügung stellen, erreicht werden.<sup>9</sup> Mit dieser Forderung legt die DFG auch gleichzeitig die Voraussetzungen für Virtuelle Forschungsumgebungen fest, die mit solchen

7 Lipp (2009, S. 501) weist in diesem Kontext darauf hin, dass Primärforschungsdaten untrennbar mit Virtuellen Forschungsumgebungen verbunden sind, da sich der Zweck solcher Umgebungen aus der Bearbeitung dieser Daten ableitet. Dies erfordert sowohl eine Infrastruktur für ihre Bereitstellung als auch die Digitalisierung analoger Forschungsdaten im Rahmen der Wissenschaftsförderung durch die DFG.

8 Im Rahmen dieses Großprojekts gilt es auch, herauszufinden, welches Design solche virtuellen Plattformen für netz-basiertes Arbeiten aufweisen müssen, um für Wissenschaftler attraktiv zu sein (vgl. Lipp, 2009, S. 503).

9 Die DFG betont dazu ausdrücklich die Relevanz der generellen Offenheit neuentwickelter Systeme für das Informationsmanagement und verweist zur Umsetzung dieser Offenheit auf Interoperabilität, offene Schnittstellen, modularen Aufbau der Systeme, die absolute Offenlegung des Quellcodes sowie die Orientierung an aktuellen Normen und Standards (vgl. Deutsche Forschungsgemeinschaft, 2011c, S. 2).



Informationssystemen und Primärdatenrepositorien interagieren sollen (vgl. Deutsche Forschungsgemeinschaft, 2011c, S. 1–4). Ein weiterer Anstoß für die Weiterentwicklung von Bibliotheken zeigt sich in der expliziten Förderung von Projekten, die sich mit der Zitierbarkeit, Langzeitverfügbarkeit und Erschließung von Primärforschungsdaten sowie der Entwicklung von Metadaten-Standards und ihrer Qualitätskontrolle auseinandersetzen. Von diesen Zielen ist es nur noch ein kleiner Schritt zum Aufbau von Umgebungen, die auch die Beforschung der Primärdaten und die Publikation der Forschungsergebnisse zu ermöglichen. In der Konsequenz dieser Förderpolitik werden Aufbau und Weiterentwicklung von Virtuellen Forschungsumgebungen als Entwicklungsfeld für Bibliotheken begriffen (vgl. Deutsche Forschungsgemeinschaft, 2010, S. 1f). Für Bibliotheken eröffnen sich große Entwicklungspotentiale dadurch, dass ihre technische und fachbezogene Kompetenz im Bereich der Informationsdienstleistungen ihnen den Wechsel vom reinen Informationsdienstleister zum Forschungspartner ermöglicht.<sup>10</sup> Diese Weiterentwicklung des ursprünglichen Aufgabenportfolios wird durch das Förderprogramm „Wissenschaftliche Literaturversorgungs- und Informationssysteme“ (LIS) der DFG vorangetrieben (vgl. Deutsche Forschungsgemeinschaft, 2006) und setzt damit einen Rahmen für die Weiterentwicklung des gesamten Bibliothekswesens. Gleichzeitig bietet dieses Programm Bibliotheken auch die Chance, Motor für neue Entwicklungen in diesem Bereich zu sein.<sup>11</sup>

In der vorliegenden Arbeit soll am Beispiel eines digitalen Korpus aus der Linguistik, dem Deutschen Textarchiv (DTA, 2011)<sup>12</sup> der Berlin-Brandenburgischen Akademie der Wissenschaften, aufgezeigt werden, welche Elemente für eine Virtuelle Forschungsumgebung in den Geisteswissenschaften notwendig und wünschenswert sind und welche Rolle Bibliotheken beim Aufbau einer Forschungsinfrastruktur spielen können. Ziel ist es, dadurch zu eruieren, welche Möglichkeiten und Potenziale sich für Bibliotheken bieten, um sich am Entwicklungs- oder Implementierungsprozess einer Virtuellen Forschungsumgebung zu beteiligen. Dabei

---

10 Einen wichtigen Schritt in diese Richtung hat die Staats- und Universitätsbibliothek Göttingen mit ihrer Teilnahme an der Virtuellen Forschungsumgebung TextGrid und der Einrichtung einer Abteilung für Forschung und Entwicklung bereits getan.

11 Der Überblick über das gesamte Programm findet sich auf den Seiten der DFG (Deutsche Forschungsgemeinschaft, 2011a).

12 Die URLs zu Projektseiten u. ä. werden zugunsten der Lesbarkeit des Textes in Kurzzitierweise angegeben und finden sich unter der entsprechenden Kurzzitation im Literaturverzeichnis wieder.

soll auch erörtert werden, welche Auswirkungen diese neue Form der Beteiligung an der deutschen Forschungslandschaft auf die Bibliotheken haben kann. Der dazu notwendige Übertritt vom tradierten bibliothekarischen Bereitstellen von Information in klassischen Containerformaten in den Bereich der Schaffung von Forschungsinfrastruktur und Forschungsmaterial ist jedoch kein einfacher und erfordert ein Umdenken in Bezug auf das Dienstleistungsportfolio der Institution Bibliothek an sich und das Berufsbild des wissenschaftlichen Bibliothekars.

## 1.1 Grids

Ian Foster und Carl Kesselman definieren *Grid* zunächst als eine Infrastruktur aus Hard- und Software, die zuverlässigen, dauerhaften, überall vorhandenen und kostengünstigen Zugang zu hochwertigen Rechenressourcen bietet Foster et al., (1999, S. 18).<sup>13</sup> Diese Definition erweitern sie später, indem sie das koordinierte Lösen von Problemen mittels Ressourcenteilung in dynamischen virtuellen Organisationen, die aus Wissenschaftlern verschiedener Institutionen auf der ganzen Welt bestehen, als grundlegende Eigenschaft von Grids angeben (vgl. ebd.).<sup>14</sup> Foster et al. (2001, S. 2) wollen darunter also mehr verstanden wissen als Datenaustausch und Arbeiten mit verteilten Rechnerkapazitäten, sondern stellen die akteurbezogene Komponente des Grids heraus:

„The sharing that we are concerned with is not primarily file exchange but rather direct access to computers, software, data, and other resources, as is required by a range of collaborative problem-solving and resource brokering strategies emerging in industry, science, and engineering. This sharing is, necessarily, highly controlled, with resource providers and consumers defining clearly and carefully just what is shared, who is allowed to share, and the conditions under which sharing occurs. A set of individuals and/or institutions defined by such sharing rules form what we call a virtual organization (VO).“

Ziel ist der kooperative soziale Zugang zu hochwertigen Rechenressourcen zum koordinierten Lösen von Problemen in dynamischen Virtuellen Organisationen (vgl. Foster et al., 1999, S. 18; Foster et al., 2001, S. 2). Der Terminus *Grid* wurde Mitte

---

13 Butler (2003) bietet eine einführende Erklärung über die Anfänge der Grid-Technologie und den speziellen Grid-Prototypen, der für die exzessiven Datenmengen des Large Hadron Collider (LHC) am Conseil Européen pour la Recherche Nucléaire (CERN) entwickelt wurde.

14 Für die Unterscheidung, ob es sich bei einer Forschungsinfrastruktur um ein Grid handelt oder nicht, hat Foster (2002) einen Katalog in Form einer Checkliste vorgelegt. Dabei betont er explizit die Relevanz offener und standardisierter Protokolle für die Interoperabilität innerhalb der Grids.

der 1990er Jahre geprägt und bezeichnete zunächst das Konzept einer verteilten Rechnerinfrastruktur (*distributed computing infrastructure*) für ressourcenintensive Berechnungen in den Natur- und Ingenieurwissenschaften (vgl. Foster et al., 1999). Jedoch verweisen Foster et al. (2001, S. 1f) ausdrücklich darauf, dass Grid-Technologien aufgrund ihres Aufgabengebiets deutlich von aktuellen Technologietrends wie Internet, Distributed- oder Peer-to-peer-Computing unterschieden werden müssen. Die konstituierende Funktion des Grids liegt im kollaborativem Lösen von Problemen mittels verteilter Ressourcen in den Virtuellen Organisationen, das entsprechend komplexe Strategien für die Verwaltung des Zugriffs auf Rechner, Software, Daten und Ressourcen erfordert. So ergeben sich aus dieser sozialen Funktion die technischen Anforderungen an die Grids.

Grids existieren als kombinierte Hard- und Software-Infrastrukturen in Netzwerken und sind somit Architekturen, die in höchstem Maße von Protokollen zur Regelung allen Datenaustauschs angewiesen sind. Als solche erfordern sie die Interoperabilität der verwendeten Protokolle, um Entwicklung, Austausch und Management von Ressourcen für Virtuelle Organisationen leisten zu können. Diese Interoperabilität muss auf (offenen) Standards und Architekturen basieren, um Erweiterbarkeit, Portabilität und Code-Sharing zu ermöglichen. Interoperabilität ist relevant, da die Partizipanten virtueller Organisationen sich sonst in ausschließlich bilateral konzipierte Austauschbeziehungen begeben müssten, ohne die Garantie, dass dieser Austausch auch mit weiteren Teilnehmern der Organisation möglich wäre.<sup>15</sup> Ist diese Multilateralität des Austauschs nicht garantiert, ist das für Virtuelle Organisationen konstituierende Element der Kollaboration nicht gegeben. Foster et al. (2001, S. 2) weisen in ihrer Definition des Konzepts der Virtuellen Organisation außerdem deutlich auf die zwingende Notwendigkeit eines hochgradig differenzierten Rechtemangements hin, ohne das Virtuelle Organisationen und damit kollaboratives Arbeiten nicht möglich sind. Daraus leiten sich als notwendige Eigenschaften für Grids Interoperabilität durch die Berücksichtigung Offener Standards und Vertrauenswürdigkeit durch differenziertes Rechtemangement ab.

---

<sup>15</sup> Wie das universelle Internetprotokoll HTTP und die HTML-Syntax gezeigt haben, ist Interoperabilität für jeglichen Datenaustausch unverzichtbar. Gleichmaßen hängen Dienste, Zugriffsberechtigungen und Ressourcenverteilung (im Sinne von Prozessorzeit und -auslastung) in Grids von Universalprotokollen ab, so dass Interoperabilität auch auf dieser Ebene unumgänglich ist (vgl. Foster et al., 2001, S. 5f).

Typologisch lassen sich Grids anhand der Anwendungsszenarien, für die sie erstellt wurden, unterscheiden (vgl. Aschenbrenner et al., 2007, S. 12): aus den Anforderungen des Zugangs zu Datenmengen im Petabyte-Bereich stammen die sogenannten *Datengrids*, die es vor allem den naturwissenschaftlichen Fächern ermöglichen, Rohdaten von Gen-Sequenzierern oder Teilchenbeschleunigern wie dem Large Hadron Collider (LHC) abzurufen und selbst zu analysieren. Häufig sind die Rechnerkapazitäten einer einzigen Einrichtung für Analysen und Simulationen mit Daten diesen Umfangs zu gering, so dass sie nur durch die kumulierte Rechenleistung eines über ein sogenanntes *Computational Grid* erzeugten virtuellen Supercomputers zu bewältigen sind. *Access Grids* hingegen bieten vor allem Kommunikationssoftware z.B. für Videokonferenzen innerhalb virtueller Organisationen, group-to-group-Kommunikation und allgemeinen Dateiaustausch.<sup>16</sup> Als generelle Gemeinsamkeit solcher Grids kann gelten, dass sie „[...]Menschen mit Hilfe von Computerressourcen in Virtuellen Organisationen zusammenbringen und das erforderliche Werkzeug für die Arbeit zur Verfügung stellen“ (Aschenbrenner et al., 2007, S. 12).<sup>17</sup> Es zeigt sich, dass Virtuelle Organisationen im Sinne Fosters erst durch Grids entstehen, indem sie Wissenschaftlern neue Möglichkeiten zum globalen, kollaborativen Arbeiten bieten und zumindest in den Natur- und Ingenieurwissenschaften die Bildung von Arbeits- und Kommunikationsgruppen ermöglichen, die über den Globus verteilt an gemeinsamen Projekten mit ebenso verteilten Daten arbeiten.<sup>18</sup>

## 1.2 Virtuelle Forschungsumgebungen

Notwendig für die Diskussion der Rolle, die sich für Bibliotheken aus der Implementierung Virtueller Forschungsumgebungen in den Forschungsprozess ergibt, ist die Klärung der Frage, was als Virtuelle Forschungsumgebung bezeichnet

---

16 Access Grids finden beispielsweise in den darstellenden Künsten Anwendung, indem über „grid-basierte Videokollaboration[en]“ (vgl. Aschenbrenner et al., 2007, S. 16) musikalische Kompositionen oder Performances verteilt entwickelt werden.

17 Zum inhaltlich verwandten, in den USA gebräuchlichen Begriff *Cyberinfrastructure* (vgl. American Council of Learned Societies, 2006, S. 6–8) und (Nentwich, 2003).

18 Zu den Traditionen und Besonderheiten der Kollaboration in den Naturwissenschaften vgl. (Knorr-Cetina, 2002, S. 227–330).

werden kann und welche Grundvoraussetzungen erfüllt sein müssen, um eine solche Umgebung bieten zu können. In der Literatur finden sich verschiedene Erklärungen des Konzepts Virtuelle Forschungsumgebung, die jedoch nicht unbedingt den Charakter einer fixierten Definition tragen. Der Begriff der Virtuellen Forschungsumgebung entwickelt sich ab etwa 2005 und geht zu diesem Zeitpunkt noch von (gridbasierten) virtuellen Lernumgebungen aus (vgl. Carr et al., 2004). Im weiteren Verlauf der Entwicklung kommt die Idee auf, dass Umgebungen dieser Art den ganzen Forschungsprozess unterstützen könnten (vgl. Wills et al., 2005). Eine erste beschreibende Definition liefert Fraser (2005), der Virtuelle Forschungsumgebungen vor allem als Teil einer erweiterbaren technischen Infrastruktur versteht, in die Werkzeuge, Dienste und Ressourcen integriert werden können.<sup>19</sup>

„A VRE [virtual research environment, J.E.K.] is best viewed as a framework into which tools, services and resources can be plugged. VREs are part of infrastructure, albeit in digital form, rather than a free-standing product. [...] The VRE, for its part, is the result of joining together new and existing components to support as much of the research processes appropriate for any given activity or role.“

Unter Ressourcen werden nach der auf Fraser (2005) basierenden Definition, die Schirnbacher (2009, S. 28) verwendet, in diesem Zusammenhang sowohl die Primärforschungsdaten und die Ergebnisse der Forschung als auch die personellen Kapazitäten im Sinne der Forscher selbst und des Servicepersonals verstanden. Die Werkzeuge bilden einzelne Teilbereiche des Forschungsprozesses oder gleich seine Gesamtheit innerhalb des mehrstufigen Frameworks ab.<sup>20</sup> Aschenbrenner et al. definieren Virtuelle Forschungsumgebungen als „formelle Infrastrukturen, in der verteilte Ressourcen und Werkzeuge vereinigt werden, um einem speziellen Zweck oder einer speziellen Nutzergruppe zu dienen“ (Aschenbrenner et al., 2007, S. 13). Neuroth et al. (2007) schließen auch Dienste in ihr mehrschichtiges Modell ein. Blanke/Hedges (2011, S. 1) bezeichnen sie in etwa als kollaborative digitale

---

<sup>19</sup> Ein Beispiel für die Erweiterung von Infrastrukturen durch Werkzeuge zur Bearbeitung verschiedener Daten ist etwa das deutsche Kooperationsprojekt European Cultural Heritage Online (ECHO). Die für dieses Projekt geschaffene Arbeitsoberfläche wird noch als scholarly workbench bezeichnet (vgl. Rieger et al., 2007).

<sup>20</sup> Auch Schirnbacher (2009, S. 28) fordert für die interdisziplinäre Nutzbarkeit von Infrastruktur und Werkzeugen deren Orientierung an offenen Standards.

Umgebungen, die die Integration von Informationsressourcen und Werkzeugen ermöglichen, um ein bestimmtes Set von Forschungsaktivitäten zu unterstützen.<sup>21</sup>

All diese Definitionen des Konzepts der Virtuellen Forschungsumgebung heben das technische Framework (Fraser, 2005; Schirnbacher, 2009), das auch als „formelle Infrastruktur“ (Aschenbrenner et al., 2007, S. 13) oder „Basisinfrastruktur“ (Neuroth et al., 2007, S. 272) bezeichnet wird, die Integration von Informationsressourcen und Werkzeugen sowie die Funktion der Unterstützung von Forschungsaktivitäten als Basiselemente hervor. Der technische Aufbau Virtueller Forschungsumgebungen ist nach Neuroth et al. (2007, S. 272–274) – unabhängig von der Disziplin, deren Forschung sie unterstützen soll – in drei Schichten unterteilbar: die technische Basisinfrastruktur, die Forschungsinfrastruktur und die daran anknüpfende Forschung selbst. Ob diese Generizität beim technischen Aufbau einer VRE enden muss, ist diskussionsbedürftig. Denkbar wäre auch, die Generizität bis auf die Ebene der Werkzeuge und Ressourcen auszudehnen, da viele Tools und Datenreservoirs von mehr als einer Disziplin oder sogar über das einzelne Wissenschaftsparadigma hinaus anwendbar sind.<sup>22</sup>

Zur Basisinfrastruktur gehört für (Neuroth et al., 2007, S. 272–274) ein Grid, das Datenzugang, Super-Computing und globale Kommunikation innerhalb Virtueller Organisationen ermöglicht, aber oft auch Kerndienste wie Authentifizierung, Autorisierung oder Accounting bietet. Die darauf aufsetzende Forschungsinfrastruktur, bietet z.B. Standards, Werkzeuge, Dienste sowie Trainingsmöglichkeiten und bildet die Schnittstelle zwischen der technologischen Basis und den Wissenschaftlern. Darüber hinaus können hier auch „höherwertige Dienste“ (Neuroth et al., 2007, S. 274) wie Langzeitarchive oder persistent identifiziert implementiert werden. Sowohl das Joint Information Systems Committee (JISC) als auch die Deutsche Forschungsgemeinschaft (DFG) heben in ihrer Definition Virtueller Forschungsumgebungen deren Rolle für den Forschungsprozess hervor (vgl. Reimer et al., 2010, S. 12–15). So definiert das JISC Virtuelle Forschungsumgebung etwa von ihrer Funktion für die Forschung her:

---

21 Was mit dem Begriff der kollaborativen digitalen Umgebungen (*digital environments*) genau gemeint ist, bleibt jedoch unklar.

22 Dieses Konzept umfassender Generizität wird in Kapitel 4 erneut aufgegriffen.

The purpose of a Virtual Research Environment (VRE) is to help researchers from all disciplines to work collaboratively by managing the increasingly complex range of tasks involved in carrying out research on both small and large scales. [...] The term VRE is now best thought of as shorthand for the tools and technologies needed by researchers to do their research, interact with other researchers (who may come from different disciplines, institutions or even countries) and to make use of resources and technical infrastructures available both locally and nationally. The term VRE also incorporates the context in which those tools and technologies are used.<sup>23</sup>

Allgemeiner noch erklärt die DFG eine Virtuelle Forschungsumgebung als Plattform für netzbasiertes kollaboratives Arbeiten, das neue Möglichkeiten für gemeinsame Kollaboration und den Umgang mit Forschungsdaten sowie Informationen eröffnet (vgl. Deutsche Forschungsgemeinschaft, 2011d).

Aus diesen verschiedenen Definitionen ergibt sich für Virtuelle Forschungsumgebungen, dass sie in erster Linie als Interface der Forscher zu einer technischen Forschungsinfrastruktur, die Forschung produktiv unterstützen kann, zu betrachten sind. Umfassende Virtuelle Forschungsumgebungen können Wissenschaftlern gleichzeitig den Zugang zu Informationsressourcen, den Werkzeugen und Diensten zu deren Beforschung und Kommunikationsmöglichkeiten mit anderen Forschern bieten. Diese erweiterten Möglichkeiten beruhen auf einer technischen Infrastruktur, die es wissenschaftsmethodisch zu nutzen gilt. Diese Ansicht unterstützen auch die Aussagen einiger Befragten der JISC Landscape Study, die 2009 durchgeführt wurde mit dem Ziel, die internationale Entwicklung von Virtuellen Forschungsgemeinschaften (*virtual research communities*) zu untersuchen und im Verhältnis zum JISC-Förderprogramm für Virtuelle Forschungsumgebungen zu evaluieren (vgl. Reimer et al., 2010, S. 5). Die Studienteilnehmer betrachten Virtuelle Forschungsumgebungen zum Einen als besagtes Infrastruktur-Interface, zum Anderen als Instrument für das Projektmanagement von Forschungsvorhaben, das sich auch in die vorhandene physische Forschungsumgebung wie Labore, Grabungsstätten und Bibliotheken integrieren muss. Dabei wird jedoch auch hervorgehoben, dass Virtuelle Forschungsumgebungen mit zunehmender Integration in den Forschungsprozess und seine soziale Umgebung die Markierung als *virtuell* verlieren und stattdessen als natürlicher Teil einer hybriden Arbeitsumgebung wahrgenommen werden (vgl. Reimer et al., 2010, S. 14f). Aus diesen Aussagen heraus postulieren Reimer et al. (2010) die Begründung des Interesses

---

<sup>23</sup> Vgl. Joint Information Systems Committee, 2009-2011

wissenschaftlicher Bibliotheken an Virtuellen Forschungsumgebungen: Bibliotheken seien bereits seit Jahrhunderten Umgebungen, in denen Forschung stattfindet. Entsprechend sei es besonders für die moderne hybride Bibliothek naheliegend, sich dafür zu interessieren, ihre Forschungsumgebung mit diesen neuen technischen Möglichkeiten auszustatten. Speziell für die Geisteswissenschaften lässt sich die Situation in diesem Zusammenhang sehr klar formulieren: Bibliotheken sind seit jeher die Labore der Geisteswissenschaften, die mit ihren Beständen unüberschaubare Mengen von Forschungsdaten anbieten. In Zeiten elektronischer Medien und der daraus entwickelten hybriden Bibliothek ist es also nur folgerichtig, den Forschern in diesen Laboren passende Arbeitsgeräte und Zugänge zu diesen neuen Medien zur Verfügung zu stellen. Unter diesem Gesichtspunkt sind Virtuelle Forschungsumgebungen die neuen Laborplätze der Geisteswissenschaften. Je stärker sich Forschung und elektronische Forschungsumgebung verzahnen, desto stärker tritt die Notwendigkeit der spezifischen Benennung dieser Forschungsumgebungen in den Hintergrund. Möglicherweise werden sich die verschiedenen Benennungen für Virtuelle Forschungsumgebungen sogar gänzlich auflösen, sobald sie in allen Disziplinen vollständig in die Forschungsarbeit integriert und alltäglich geworden sind (vgl. Reimer et al., 2010, S. 15). Für eine gültige Definition des Terminus *Virtuelle Forschungsumgebung* lässt sich abschließend festhalten, dass sie sowohl dem technischen Hintergrund des Konzepts als auch seiner Funktion für den Forschungsprozess Rechnungen tragen muss. Im folgenden soll in Anlehnung an die dargelegten Definitionen des JISC, der DFG sowie Fraser (2005) und Schirmbacher (2009) folgende Definition gelten:

Eine Virtuelle Forschungsumgebung ist eine mehrstufige, netzbasierte Infrastruktur aus Soft- und Hardware, die Wissenschaftlern die Forschungsarbeit durch die teilweise oder vollständige Abbildung des Forschungsprozesses in Form von Ressourcen, Werkzeugen und Diensten ermöglicht oder erleichtert.

## **2. Digital Humanities – Geisteswissenschaft im digitalen Raum**

Die neuen Methoden, die der Wissenschaft durch entsprechende technische Infrastruktur wie Grids und Virtuelle Forschungsumgebungen zur Verfügung stehen, wirken nicht nur auf das Forschungsprozedere an sich, sondern auf die gesamte



Wissenschaftskultur zurück. Allen voran ist das naturwissenschaftliche Paradigma dadurch insgesamt zu stärkerer Vernetzung der Forschungsaktivitäten und engerer Kollaboration übergegangen. Die erweiterten Möglichkeiten digitalen wissenschaftlichen Arbeitens erlauben es Wissenschaftlern, neue Arten von Publikationen wie etwa Datenbanken oder XML-Korpora zu erzeugen, die wiederum zu komplexeren sogenannten *Forschungsobjekten*<sup>24</sup> zusammengesetzt und wiederum beforscht werden können (vgl. Blanke et al., 2011, S. 2). Diese neue Art des Forschens wird in Deutschland und Großbritannien unter dem Begriff *eScience*, im Sinne von *electronic science* oder *enhanced science*, diskutiert und gefördert.<sup>25</sup> In zunehmendem Maße wenden sich auch die Geisteswissenschaften diesem neuen Forschungsparadigma zu und adaptieren die darunter angegebenen Techniken und Konzepte. Entsprechend wird mit den Begriffen *eHumanities* und *Digital Humanities* der Bereich des Wissenschaftskontinuums benannt, „an dem neue Arbeitsweisen und Technologien in die Geistes-, Kultur und Sozialwissenschaften hineingetragen werden und umgekehrt der eigene Beitrag zu den bestehenden e-Science-Aktivitäten der Naturwissenschaften gebündelt wird“ (Neuroth et al., 2007, S. 273).<sup>26</sup>

Der Nutzen der erweiterten Forschungskapazitäten, die Grids und Virtuelle Forschungsumgebungen bieten, liegt für die Geisteswissenschaften vor allem in der Vernetzung projektspezifischer Daten (z.B. Thesauri, Lexika, Namenslisten etc.) und verteilter Fachkenntnisse sowie den dazugehörigen Werkzeugen (vgl. Aschenbrenner et al., 2007). Als paradigmatisches Projekt für die Weiterentwicklung der eHumanities erweist sich die Virtuelle Forschungsumgebung, die für die archäologische Erforschung der Silchester-Ausgrabungsstätte in Großbritannien entwickelt wurde und beispielsweise direkt vor Ort digitalisierte Daten in Echtzeit von

---

24 Forschungsobjekte dieser Art entstehen aus einer „Vielzahl von Informationsobjekten [,die ] mit einander in Wechselbeziehung stehen“ (Blanke et al., 2011, S. 2). Insofern weist der Begriff des *Forschungsobjekts* in dieser Intension Parallelen mit dem der *Entität* auf. Damit wird die Extension des Begriffs des *Forschungsobjekts* über den des *Forschungsgegenstands* hinaus erweitert. Das bedeutet gleichzeitig eine Erweiterung der Perspektive dessen, was alles Gegenstand der Forschung sein kann und wie dieser Gegenstand modelliert sein muss und eine Veränderung der ontologischen Grundlagen.

25 Einen Überblick gibt Schreibman (2011).

26 Einen institutionellen Unterstützungsansatz für die eHumanities bietet das Forschungsnetzwerk Digital Research Infrastructure for the Arts and Humanities (DARIAH), dessen Ziel es ist, eine umfassende digitale Forschungsinfrastruktur für die Geistes- und Gesellschaftswissenschaften zu schaffen und auf dieser interdisziplinären Austausch von Forschungsdaten, -methoden und -kenntnissen zu ermöglichen (vgl. DARIAH, 2012).

der Grabungsstätte in eine Projektdatenbank überträgt und sofortigen Zugriff durch abwesende Forscher ermöglicht (vgl. Joint Information Systems Committee, 2009; Aschenbrenner et al., 2007, S. 13). Das erste deutsche eHumanities-Projekt ist die Virtuelle Forschungsumgebung TextGrid (TextGrid, 2012), die durch ein deutschlandweites Netzwerk verschiedener universitärer Forschungspartner unter Leitung der Staats- und Universitätsbibliothek Göttingen entwickelt wird (vgl. Küster et al., 2009, S. 185).<sup>27</sup> Virtuelle Forschungsumgebungen bieten den Geisteswissenschaften bereits durch ihre mehrschichtige Struktur verschiedene Vorteile: Erstens bietet die integrierte Unterstützung des elektronischen Publizierens Möglichkeiten der Förderung wissenschaftlichen Nachwuchses durch die Unabhängigkeit von etablierten Kanälen mit hohen Eintrittshürden. Zweitens werden über die Forschungsinfrastruktur Dienste wie Datenverwaltung, Suche oder automatische Textanalyse zur Verfügung gestellt. Drittens kann die Forschung selbst schon bereits durch die Anforderungen, die sie an ihre Forschungsinfrastrukturen stellt, einen Beitrag zur Weiterentwicklung nützlicher Dienste und Werkzeuge leisten und dabei gemeinsam mit Infrastrukturanbietern Standards definieren, die Interoperabilität und semantische Vernetzung von Inhalten, Daten und Werkzeugen ermöglichen. Gleichzeitig gibt die Infrastruktur der Forschung Dienste und Basistechnologien an die Hand, die das Wissenschafts- und Projektmanagement vereinfachen, indem sie Authentifizierungs-, Autorisierungs- und Abrechnungsdienste sowie Speicher, Rechenleistung und Gelegenheiten zur Langzeitarchivierung bieten. Die Voraussetzung für eine gute Akzeptanz dieser Techniken seitens der Forschung liegt jedoch im *plug-and-play*-Charakter dieser Werkzeuge, die auf Anhieb und ohne großen Einarbeitungsaufwand für die Wissenschaftler funktionieren müssen. Zuletzt bergen Virtuelle Forschungsumgebungen den Geisteswissenschaften das Potenzial zur Globalisierung, das in den Naturwissenschaften schon seit einiger Zeit genutzt wird (vgl. Neuroth et al., 2007, S. 274–276). Projekte wie der Virtuelle Raum Reichsrecht (ViRR) des Max-Planck-Instituts für Rechtsgeschichte (Virtuelle Raum Reichsrecht 2010), eAQUA (eAQUA, 2012), Monumenta Germanica Historiae digital

<sup>27</sup> Projektpartner des TextGrid-Netzwerks sind die Technische Universität Darmstadt, das Institut für Deutsche Sprache (Mannheim), das Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften (Universität Trier), die Fachhochschule Worms, das Kompetenzzentrum für EDV-Philologie der Universität Würzburg, DAASI international GmbH, Saphor GmbH und die erwähnte Staats- und Universitätsbibliothek Göttingen. Ein eingehender Überblick bieten die Projektseiten auf (TextGrid, 2012).

(Monumenta Germaniae Historica, 2011), Nietzsche Source (Nietzsche Source, 2011), Perseus (Perseus Digital Library weiterführend dazu Smith et al., 2000), das Heinrich-Heine-Portal der Universität Trier (Heinrich-Heine-Portal, 2011), Histodillo (Histodillo, 2012), Clio Online (Clio-online, 2012; Thaller, 1980) oder H-Soz-u-Kult (H-Soz-u-Kult, 2012; Meyer, 2011) sind nur einige Beispiele für eHumanities-Aktivitäten, die im Kontext Virtueller Forschungsumgebungen Daten bieten und teils national, teils international vernetzen, Data-Mining-Technologien für die Geisteswissenschaften nutzbar machen und Werkzeuge für digitale Forschung zur Verfügung stellen.<sup>28</sup> Ob und wie dieses Potenzial genutzt wird und ob Grids die einzig sinnvolle Möglichkeit darstellen, Daten für die Geisteswissenschaften komfortabel zugänglich zu machen, bleibt im folgenden zu diskutieren.<sup>29</sup>

## **2.1 Grundlagen wissenschaftlichen Arbeitens in den Geisteswissenschaften**

Um Virtuelle Forschungsumgebungen für die Geisteswissenschaften wirklich effizient nutzbar zu machen, bedarf es einer Reflexion der wissenschaftstheoretischen Kernkonzepte dieses Paradigmas. Erst wenn diese identifiziert und innerhalb einer Virtuellen Forschungsumgebung abgebildet sind, kann eine Anpassung der Dienste und Werkzeuge, die auf der Infrastruktur aufsetzen, für diese Konzepte vorgenommen werden. Ziel ist es, wissenschaftliche Primitive herauszuarbeiten, die das Arbeiten in virtuellen Forschungsinfrastrukturen abbilden und so dazu beitragen, die Infrastruktur derart weiterzuentwickeln, dass sie die Grundaktivitäten aller Wissenschaftsparadigmen befriedigen kann. Dadurch müsste nicht mehr jede Einzeldisziplin eine speziell auf ihre Bedürfnisse zugeschnittene Virtuelle Forschungsumgebung haben. Stattdessen kann bereitgestellte Basis- und Forschungsinfrastruktur, die wissenschaftliche Kernaktivitäten unterstützt, als Grundgerüst dienen, das durch zusätzliche fachspezifische Module erweitert und damit auf disziplinspezifische Bedürfnisse angepasst wird. Um diese Anpassungsfähigkeit zu gewährleisten, muss jedoch zuerst ein generisches Rahmenwerk geschaffen werden, das spezielle Werkzeuge und Module aufnimmt. Da Virtuelle Forschungsumgebungen nach Blanke et al. (2011, S. 1) ein breites

---

<sup>28</sup> Einen Überblick über die genannten Beispiele bieten Neuroth et al. (2009).

<sup>29</sup> Siehe Kapitel 2.2 zu technischen Anforderungen.

Konzept darstellen und daher ein generisches Konstrukt sein sollen, ist die Berücksichtigung der wissenschaftlichen Primitive des jeweils zu bedienenden Forschungsparadigmas (d.h. geistes- oder naturwissenschaftliches Paradigma) die Voraussetzung für die Entwicklung generischer, multidisziplinär nutzbarer Virtueller Forschungsumgebungen, die mehr als nur einer einzigen Disziplin innerhalb dieses Paradigmas nutzen können Blanke et al. (2011, S. 2). Dazu müssen generische Infrastrukturen aufgebaut werden, die wissenschaftliche Primitive ganzer Wissenschaftsparadigmen, z.B. der Geisteswissenschaften, abdecken und erst in zweiter Instanz flexibel für Einzeldisziplinen wie etwa die Linguistik durch Werkzeuge und Dienste angepasst werden. Diese Form der Generizität sei jedoch ohne die Orientierung an Offenen Standards nicht denkbar (vgl. ebd.). Vor der Frage, welche Anforderungen eine Disziplin oder eine Forschungsgemeinschaft an Virtuelle Forschungsumgebungen stellt, muss also die Frage stehen, auf welchen wissenschaftlichen Grundaktivitäten das jeweilige Wissenschaftsparadigma fußt. Virtuelle Forschungsumgebungen können nur dann flexibel einsetzbar und dennoch ausreichend forschungsspezifisch sein, wenn bei ihrer Entwicklung und Implementierung in das wissenschaftliche Arbeiten klar ist, welche Grundbedürfnisse wissenschaftlichen Arbeitens sie unterstützen sollen. Verschiedene Projekte und Studien haben sich daher in den vergangenen Jahren zum Ziel gesetzt, zu analysieren, welche wissenschaftlichen Kernaktivitäten den verschiedenen Wissenschaftszweigen zugrundeliegen (u.a. Egan, 1954; Gould, 1988; Gould et al., 1989; Gould et al., 1991; Bakewell et al., 1988). Besonderes Augenmerk muss in der heutigen Situation vor allem auf Kategorisierungsansätze gelegt werden, die die Besonderheiten des Forschens im digitalen Raum berücksichtigen. In jüngerer Zeit hat sich Unsworth (2000) vor allem auch vor dem Hintergrund des Einflusses moderner Technologien auf den Forschungsprozess auseinandergesetzt.<sup>30</sup> Die Technologien und die damit verbundene Arbeit mit hybriden Ressourcen (d.h. digitale und analoge Dokumente) charakterisiert sich für ihn als kollaborativ, multimedial, multidisziplinär und multitechnologisch sowie durch ihre groß und langfristig angelegten, komplexen Forschungsvorhaben (vgl. Short, 2005, S. 20). Er extrahiert sieben Basisaktivitäten wissenschaftlichen Arbeitens, die er über alle

---

<sup>30</sup> Unsworth (2000) entwickelt seine wissenschaftlichen Primitive ausgehend von aristotelischer Theorie und Terminologie.

Disziplinen und Zeiten hinweg und als unabhängig von der theoretischen Orientierung der Forschung ansieht und bezeichnet sie erstmals als sogenannte *wissenschaftliche Primitive* (Unsworth, 2000, Übersetzung J.E.K.)<sup>31</sup>: *Entdecken (discovering)*, *Annotieren (annotating)*, *Vergleichen (comparing)*, *Referenzieren (referring)*, *Erproben (sampling)*, *Erklären (illustrating)*, *Darstellen (representing)*. Diese wissenschaftlichen Primitive nehmen Palmer et al. (2009) auf und unterziehen sie innerhalb ihres Projekts einer Revision (vgl. Tab. 1, S. 23).<sup>32</sup> Aus einem Vergleich von drei Forschungsberichten über zwanzig Disziplinen extrahierten sie *Suchen (searching)*, *Sammeln (collecting)*, *Lesen (reading)*, *Schreiben (writing)*, *Zusammenarbeiten (collaborating)* (vgl. Palmer et al., 2009, Übersetzung J.E.K.) als disziplinübergreifene wissenschaftliche Primitive, die sich ihrerseits wiederum in verschiedene feiner granulいたe Teilaktivitäten aufgliedern (vgl. Palmer et al., 2008; Brockman et al., 2001; Palmer et al., 2002).<sup>33</sup>

Abbildung 1 (S. 22) gibt einen Überblick, wie wissenschaftliche Primitive bzw. Aktivitäten nach Palmer et al. (2009) auf die beiden großen Paradigmen Naturwissenschaft und Geisteswissenschaft und ihre interdisziplinären Verknüpfungen verteilt sind.<sup>34</sup>

---

31 Die Übersetzung *wissenschaftliche Primitive* beschreibt den basalen Charakter der auf den Forschungsprozess bezogenen Handlungen und wird daher - ungeachtet eventueller weiterer Übersetzungen - als deutsche Übersetzung des Terminus *scholarly primitives* verwendet werden.

32 Palmer et al. (2009, S. 7) entfernen sich dabei bewusst vom Begriff des wissenschaftlichen Primitivs zugunsten des Konzepts der *scholarly information activities*, das die Rolle der Informationsarbeit in diesem Kontext hervorheben soll (vgl. Palmer et al., 2008).

33 Wie Blanke et al. (2011, S. 2) mit Verweis auf Unsworth (2000) und Bernardou et al. (2010) anmerken, ist die Klassifikation von Palmer et al. (2009) nur eine unter vielen, stellt aber einen geeigneten Konzeptionsrahmen für weitere Analysen dar.

34 Für Palmer et al., 2009, S. 35) gehören die Sozialwissenschaften je nach Forschungsgegenstand als historische Sozialwissenschaften zum geisteswissenschaftlichen bzw. als quantitative Sozialwissenschaften zum naturwissenschaftlichen Paradigma (vgl. Palmer et al., 2009, S. 35).

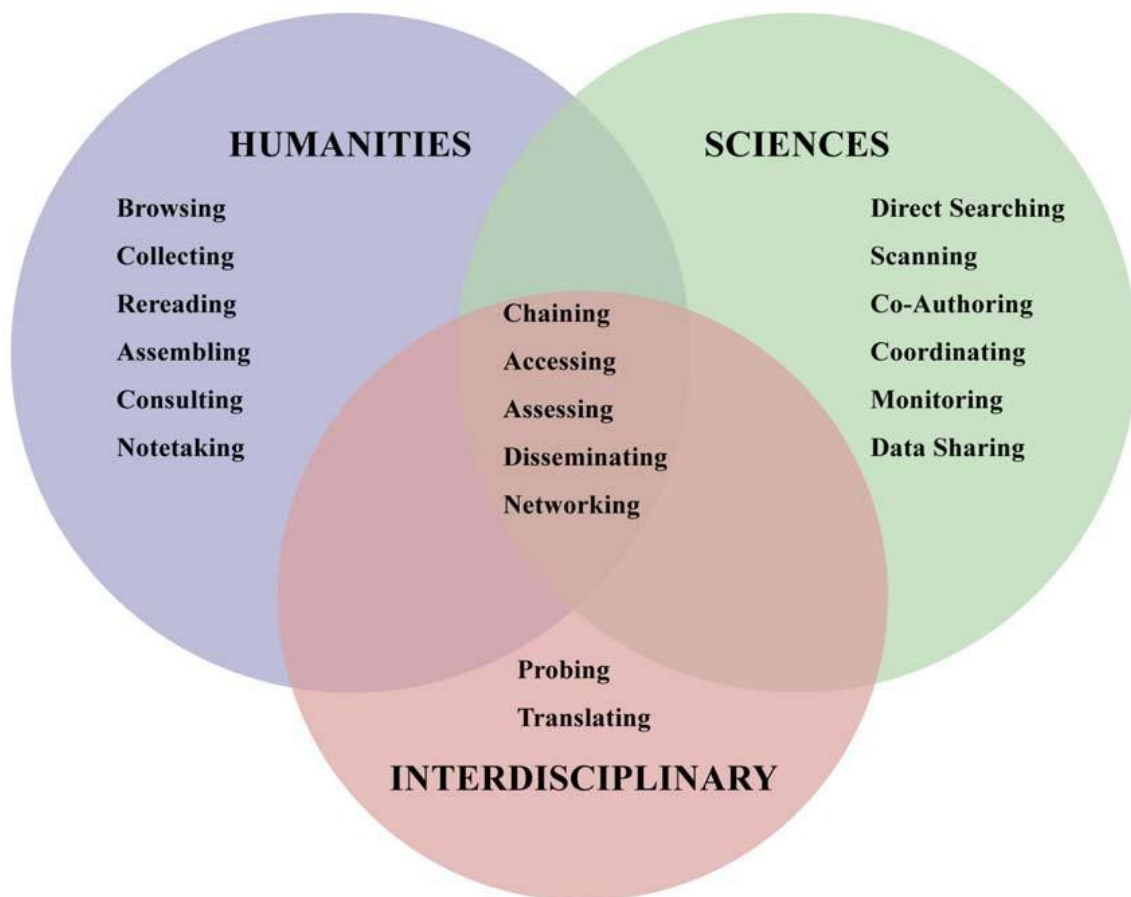


Abbildung 1: Zuordnung wissenschaftlicher Primitive zu Wissenschaftsparadigmen nach Palmer et al. (2009: 35)

Blanke et al. (2011) betrachten diese Primitive als konzeptionellen Rahmen für die Einordnung wissenschaftlicher Aktivitäten und verwenden sie deshalb als Grundlage für die Konzeption und Entwicklung von Forschungsinfrastruktur, die diese Basishandlungen unterstützen soll Blanke et al. (2011, S. 2). Von dieser Basis ausgehend entwickeln sie ein Modell lose gekoppelter modularer Infrastrukturkomponenten, die den Prozess der Verarbeitung von wissenschaftlichen Rohdaten, d.h. Informationsressourcen, zu diskutierbaren Forschungsergebnissen vergleichbar mit einem Workflow abbilden (ebd.). Da die Klassifikation nach Palmer et al. (2009) das infrastrukturbasierte Arbeiten aber noch nicht miteinbezieht und sie ihre wissenschaftlichen Primitive stark an den textgebundenen Wissenschaften auszurichten scheinen, erarbeiten Blanke et al. (2011) basierend auf dieser Einteilung eine Klassifikation, die einerseits auch die empiriebezogenen

Gesellschaftswissenschaften in den Blick nimmt und andererseits die neuen Entwicklungen innerhalb des Forschungsprozesses, die durch die Verwendung digitaler Technologien entstehen, widerspiegeln: *Entdecken (discovering)*, *Sammeln (collecting)*, *Vergleichen (comparing)*, *Abliefern (delivering)* und *Zusammenarbeiten (collaborating)* (vgl. Blanke et al., 2011, S. 2, Übersetzung J.E.K.).

wissenschaftliche Primitive - (Unsworth, 2000), Übersetzungen J.E.K	wissenschaftliche Primitive - (Palmer et al., 2009), Übersetzung J.E.K.	wissenschaftliche Primitive -(Blanke et al., 2011), Übersetzung J.E.K.
Entdecken (discovering)	Suchen (searching)	Entdecken (discovering)
Annotieren (annotating)	Sammeln (collecting)	Sammeln (collecting),
Vergleichen (comparing)	Lesen (reading)	Vergleichen (comparing)
Referenzieren (referring)	Schreiben (writing)	Abliefern (delivering)
Erproben (sampling)	Zusammenarbeiten (collaborating)	Zusammenarbeiten (collaborating)
Erklären (illustrating)		
Darstellen (representing)		

Tabelle 1: wissenschaftliche Primitive nach Unsworth (2000), Palmer et al. (2009) und Blanke/Hedges (2011) im Vergleich

Insgesamt weisen die je fünf wissenschaftlichen Primitive, die von beiden Studien angegeben werden, dennoch große Ähnlichkeit auf (vgl. Tab. 1, S. 23).<sup>35</sup> Anhand der großen Ähnlichkeiten zwischen den Klassifikationen hinsichtlich der Einteilungen und Begriffe ist davon auszugehen, dass diese Primitive tatsächlich die Basisaktivitäten des Forschens im Allgemeinen widerspiegeln, auch wenn sie in verschiedenen Studien unterschiedlich differenziert werden. Nicht alle genannten Primitive sind gleichermaßen im analogen bzw. digitalen Raum beheimatet. Während die Primitive Entdecken, Sammeln und Vergleichen aus dem analogen Forschungsraum stammen und erst auf den digitalen Raum abgebildet werden müssen, sind die Primitive des (elektronischen) Abliefers und der globalen Zusammenarbeit genuin netzbasierte Prozesse, wenngleich einige ihrer Einzelkomponenten (etwa das Schreiben, das am Beginn des Ablieferungsprozesses steht) durchaus ihre Wurzeln im analogen Raum haben. Aus dieser Perspektive zerfallen die genannten Aktivitäten in die Gruppe der traditionellen Primitive (Entdecken, Sammeln, Vergleichen), die erst in den digitalen Raum übertragen werden müssen und die Gruppe der modernen Primitive (Abliefern, Zusammenarbeiten), die erst durch den digitalen Raum entstehen. Beide Gruppen müssen bei der Entwicklung von Virtuellen Forschungsumgebungen adäquate Berücksichtigung finden. Da die von Blanke et al. (2011) vorgelegte Klassifikation sich am stärksten auf das wissenschaftliche Arbeiten im digitalen Raum bezieht, soll sie als Grundlage der weiteren Ausführungen dienen.

## **2.2 Neue Arbeitsformen in den Geisteswissenschaften**

Die neuen technologischen Möglichkeiten, die sich für wissenschaftliches Arbeiten aus der Existenz von Grids, Virtuellen Forschungsumgebungen und anderen digitalen Technologien ergeben, wurden in den Natur-, Ingenieur- und Humanwissenschaften schnell angenommen und haben im Forschungsalltag der zugehörigen Disziplinen bereits eine gewisse Selbstverständlichkeit erlangt.<sup>36</sup> Die

<sup>35</sup> Das prozessbezogene 'Suchen' bei Palmer et al. (2009) korrespondiert mit dem ergebnisbezogenen 'Entdecken' bei Blanke/Hedges (2011). Das 'Schreiben' bei Palmer et al. (2009) erscheint als traditionsorientierte Art des 'Abliefers' bei Blanke/Hedges (2011). Ein Unterschied besteht jedoch zwischen 'Lesen' bei Palmer et al. (2009:19-21) und 'Vergleichen' bei Blanke/Hedges (2011: 2). Lesen beinhaltet Vergleichen und erweist sich damit als das umfassendere Konzept.

<sup>36</sup> Beispielsweise wäre die Analyse von Daten in der Bioinformatik oder Hochenergiephysik ohne Rechencluster und Grids kaum denkbar.



Disziplinen des naturwissenschaftlichen Paradigmas profitieren schon seit geraumer Zeit von den Vorzügen gridbasierter Infrastruktur für ihren Forschungsprozess: direkte Kommunikation mit weit entfernten Forschungspartnern, spontaner Austausch auch großer Datenmengen, Direktzugriff auf verteilte Rohdaten inklusive Software zu deren Verwaltung und Analyse sowie kooperatives Bearbeiten derselben Datenbasis mit gegenseitiger Zuarbeit (vgl. Neuroth et al., 2007, S. 273). So groß ihre Akzeptanz in diesen Wissenschaftszweigen ist, so reserviert gehen die Geistes- und Gesellschaftswissenschaften jedoch mit dem *Digital Turn*<sup>37</sup> um (vgl. Rapp, 2010, S. 4). Gerade der kollaborative Ansatz, der vor allem durch Access Grids und die darin integrierte Social Software gefördert wird, und die Gemeinschaftsleistung vor der Einzelleistung hervorhebt, wirkt auf das traditionsorientierte geisteswissenschaftliche Paradigma, dessen Forschungserkenntnisse in erster Linie auf Einzelleistungen beruhen, geradezu bedrohlich. Ein Paradigmenwechsel der gesamten Forschungskultur scheint gefordert (vgl. Neuroth et al., 2007, S. 272).

Eine der großen Besonderheiten des geisteswissenschaftlichen Arbeitens liegt darin, dass Geisteswissenschaftler bis heute trotz aller technologischen Veränderungen und Möglichkeiten in der Regel allein arbeiten (vgl. Stone, 1982), statt kooperatives Arbeiten zu bevorzugen, wie es in den Naturwissenschaften üblich ist (vgl. Knorr-Cetina, 2002)<sup>38</sup>. Der Grund für das Festhalten an der wissenschaftlichen Monade liegt im wissenschaftstheoretischen Modus der Generierung von Wissen in den Geisteswissenschaften begründet: die erarbeitete Meinung des einzelnen Wissenschaftlers zu einer Forschungsfrage oder einem Werk gilt immer noch als wichtigster Beitrag zum kollektiven Wissensschatz der Disziplin (vgl. Stone, 1982, S. 294f). Wissen resultiert nach Reagor/Brown (Reagor et al., 1978, S. 242f) innerhalb dieses wissenschaftlichen Paradigmas daraus, dass sich ein einzelner Geist forschend mit einem Ausschnitt der Realität beschäftigt und diesen im Kontext der eigenen Erfahrung und des eigenen Verstehens neu interpretiert. Ein reger informeller Austausch der dabei entstehenden Ergebnisse erfolgt jedoch nicht, weshalb in den Geisteswissenschaften auch Netzwerke wie die in den Naturwissenschaften typischen *invisible colleges* fehlen, so dass die Verbreitung von

---

<sup>37</sup> Der Begriff des *Digital Turn* steht in Analogie zum Empirical Cultural und Linguistic Turn, die ihrerseits auch paradigmenerändernde Prozesse in ihren Wissenschaftsfeldern bezeichnen.

<sup>38</sup> Vgl. auch American Council of Learned Societies (2006)

Wissen kaum auf Diffusion über informelle soziale Interaktion beruht, sondern in vornehmlich im öffentlichen Raum institutionalisierten Publikationen erfolgt.<sup>39</sup> Dabei liegen Materialsuche und -interpretation häufig so nahe beieinander, dass die Suche selbst bereits einen wichtigen Teil der Forschungsarbeit ausmacht, weswegen das stöbernde Suchen (*browsing*) für Forscher sehr wichtig ist (vgl. Stone, 1982, S. 294f). Daraus erklärt sich gleichzeitig, warum Literaturrecherche in der Regel nicht delegiert werden kann. Zum einen fehlt das Vertrauen in die Fähigkeiten anderer, eine evaluative Auswahl aus dem Gesamtkanon zu treffen, zum anderen bereitet es den Wissenschaftlern Schwierigkeiten, exakt zu kommunizieren, was sie suchen. Moderne Bibliotheksbestände implizieren aber, dass dem Nutzer klar ist, wonach er sucht, was bei Geisteswissenschaftlern jedoch häufig nicht der Fall ist (ebd.). Häufig wird erst bei der Durchsicht vorhandener Bestände und ihrer Verweise auf andere Titel klar, was verwendbar und zu beschaffen ist.<sup>40</sup>

Sicherlich wird die Leistung des Einzelnen in den Geisteswissenschaften auch zukünftig ihre tragende Rolle behalten. Jedoch stellt sich die Frage, ob die in den Geisteswissenschaften typische Aversion gegen kollaboratives Arbeiten möglicherweise auf einer inadäquaten Auffassung von Kollaboration nach dem naturwissenschaftlichen Modell beruht. Auch innerhalb des naturwissenschaftlichen Paradigmas ist die Leistung des einzelnen Wissenschaftlers ausschlaggebend und wird als solche honoriert – etwa ist in den meisten Fächern an der Reihenfolge der Autorennennung in Zeitschriftenaufsätzen deutlich erkennbar, wer welchen Anteil an der Gesamtleistung eingebracht hat.<sup>41</sup> Gelänge es den Geisteswissenschaften, eine Wertschätzung des kollaborativen Ansatzes zu entwickeln, könnten sie vom *Digital Turn* als erneutem Paradigmenwechsel nach dem *Empirical* und dem *Cultural Turn*

---

39 Der Begriff *invisible college* geht zurück auf Crane (1972, S. 34f), die in einer wissenssoziologischen Studie den Wissenszuwachs in den Naturwissenschaften durch Diffusionseffekte postuliert. Invisible colleges sind soziale Netzwerke in der Wissenschaft, in denen kollaborierende Arbeitsgruppen vernetzt sind, so dass durch soziale Ansteckungseffekte die Ausbreitung (Diffusion) von Wissen zustandekommt.

40 Verschiedene Typen von Arbeiten, die im Rahmen geisteswissenschaftlicher Forschung entstehen sowie dem dazugehörigen wissenschaftlichen Vorgehen erklärt Stone (1982, S. 296). In Bezug auf Forschungsmaterialien gilt, dass Monographien und Herausgeberschriften in der Regel wichtiger sind als Zeitschriften und auch alten Schriften eine sehr hohe Relevanz zukommt, da aktuelle Ansichten stets mit den tradierten verglichen werden (ebd.), denn traditionelle Bearbeitungen des Materials sind auch wichtig zur Abgrenzung der eigenen Arbeit im Vergleich zu anderen Bearbeitungen.

41 Dies erfordert jedoch die genaue Kenntnis des sozialen Kodex dieser Forschungskultur.

nur profitieren. Doch kollaboratives, interdisziplinäres Arbeiten in verteilten und vernetzten Forschungsverbänden bietet Chancen zur Bewältigung großer Forschungsvorhaben, die vorher aufgrund ihres Umfangs fast undenkbar wären und erst durch die neuen technischen Potentiale möglich werden. Der Preis für diese Erweiterung des wissenschaftlichen Möglichkeitsraums ist jedoch die Bewältigung der Herausforderung, althergebrachte soziale Muster, die der Forschungskultur zugrundeliegen, zu verändern.<sup>42</sup> Durch kollaboratives, technologisch unterstütztes Arbeiten können Forschungsprojekte durchgeführt werden, die bereits aufgrund des großen Umfangs der zu beforschenden Datengrundlage in Einzelarbeit bzw. zersplitterten Einzelprojekten innerhalb einer überschaubaren Arbeitsgruppe niemals zu bewältigen wären. Aschenbrenner et al. (2007, S. 18) formulieren dazu:

„Die nicht zu überblickende Fülle der Quellen – auch nur für Ausschnitte einer Disziplin – bringt eine fast zwangsläufige Spezialisierung und Diversifizierung der Wissenschaftsdisziplinen wie der Forscherindividuen mit sich. Dies hat auch zur Folge, dass Forschungen und Forschungsgegenstände „begrenzt“ bleiben. Geisteswissenschaftliche Forschung im Allgemeinen wie auch die textwissenschaftliche im Besonderen ist ferner in besonderer Weise an Personen gebunden, und daher hat die individuelle Einzelleistung des Forschers einen hohen Stellenwert und wird diesen auch behalten.“

Das Arbeiten im Rahmen einer Virtuellen Forschungsumgebung, die kollaborative Ansätze unterstützt, kann also Synergieeffekte erzeugen, da Forschungsfragen aus den verschiedenen Perspektiven unterschiedlicher Wissenschaften betrachtet und so multidimensional angegangen werden können. Dabei wird auch die interdisziplinäre Zusammenarbeit über global verteilte Quellen für mehrere Disziplinen möglich. Neuroth et al. (2007) verweisen in diesem Kontext auf weitere Vorteile: im Zuge multiperspektivisch ausgerichteter transdisziplinärer Zusammenarbeit entstehen neue Fragestellungen, die in kollaborativen Verbänden schneller und mit besseren Ergebnissen bearbeitet werden können bei gleichzeitiger Weiterentwicklung der Daten und Werkzeuge sowie Vereinfachung der Publikation der Ergebnisse. Ziel einer Virtuellen Forschungsumgebung ist es nach Neuroth et al. (2007, S. 273), den gesamten Arbeitsprozess so zu unterstützen, dass die Wissensgenese erleichtert und der Wissensaustausch über die Disziplinen hinweg beschleunigt wird.

---

42 Ein in diesem Zusammenhang häufig anzutreffendes Problem, das die Zusammenarbeit erheblich erschweren kann, ist die fehlende Verständigung auf eine gemeinsame Terminologie. Daher muss am Beginn einer interdisziplinären Kollaboration immer die Klärung des kollektiv genutzten Fachvokabulars stehen.

Die Nutzung dieser Vorteile erfordert jedoch in Bezug auf jedes geplante Einzelprojekt die Entwicklung einer zentralen Fragestellung, an der alle beteiligten Disziplinen gleichermaßen interessiert sind. Es muss also über das Interesse an der Nutzung von technischer Infrastruktur und ihren Werkzeugen ein Themenkonsens gefunden werden. Das Erarbeiten eines solchen sowie das kollaborative Arbeiten selbst erfordern das gegenseitige Vertrauen der Projektteilnehmer und ein gegenseitiges Verständnis ihrer jeweiligen Fachsprachen und -kulturen. Dieses Vertrauen ist vor allem nötig, da die Transparenz, die durch die Arbeit mit gemeinsamen Primärdaten entsteht, auch die Einsehbarkeit in einzelne Forschungsprojekte innerhalb des Wettbewerbsfelds und durch angrenzende Felder bedeutet und die Befürchtung steigert, Projektideen und Forschungsansätze könnten kopiert werden.

Die Umstellung der Forschungskultur von der wissenschaftlichen Monade auf kollaboratives, zumindest aber kooperatives Arbeiten bedeutet einen grundsätzlichen Perspektivenwechsel auf die Erarbeitung von Forschungsergebnissen und das 'geistige Eigentum'<sup>43</sup> daran: solche Ergebnisse sind durch die kooperative Entstehung nicht mehr intellektueller 'Alleinbesitz' einer einzigen Person, sondern müssen als Errungenschaft einer ganzen Gemeinschaft von Forschern und ihrer vereinten geistigen Kräfte angesehen werden. Die Investition des eigenen Wissens und persönlicher intellektueller Leistung darf dabei nicht als 'Weggeben' oder 'Hergeben geistigen Eigentums' angesehen werden, sondern als Beitrag, der das Projekt, alle seine Beteiligten und den Erkenntnisgewinn der Disziplin selbst voranbringt. Jede Investition eigenen Wissens bedeutet in diesem Kontext gleichzeitiges Profitieren vom Wissen der anderen Beitragenden, so dass die Inspiration zu neuen Erkenntnissen und Ideen des Einzelnen immer gespeist wird

---

43 Der Begriff *geistiges Eigentum* verbindet das Konzept des Privateigentums sprachlich auf subtile Weise mit dem Konzept des Monopols auf geistige Inhalte. Die beiden Konzepte gehören jedoch nicht zur selben Kategorie. Der Begriff *geistiges Eigentum* legt nahe, Geistiges im Sinne von Ideen und kreativem Schaffen könne Eigentum sein. Dem widersprechen jedoch zwei Argumente. Zum einen ist das Eigentumsrecht an einem Objekt nicht zeitlich begrenzt im Gegensatz zum Urheberrecht an einem Werk oder Monopolen auf Erfindungen. Zum zweiten können Ideen und ihre kreativen Inhalte nicht gestohlen werden wie ein Objekt. Es ist nicht möglich, sie durch Kopieren für den Entwickler der Idee unbenutzbar zu machen. Eine gestohlene CD könnte weder erneut angehört noch verkauft werden und fehlt dem Besitzer somit. Ein kopiertes Lied kann jedoch weiterhin gesungen und Aufnahmen davon verkauft werden. Durch die Kopie und Weiterverwendung entsteht also kein Fehlen und keine Verknappung des Inhalts (vgl. Grote, 2011, S. 22–26).

aus den vorherigen oder parallelen Investitionen der anderen Beiträger. Somit muss diese Investition auch wieder Eingang in die Gemeinschaftsanstrengung, der sie entspringt, zurückfinden. Für Aschenbrenner et al. (2007, S. 11) stellt dies einen schwer zu vollziehenden Paradigmenwechsel in der geisteswissenschaftlichen Forschungskultur dar.

Die Nutzung der neuen Möglichkeiten wissenschaftlichen Arbeitens werden nicht nur angeboten, sondern auch institutionell eingefordert, indem Förderungseinrichtungen wie die DFG oder das BMBF auch von Geisteswissenschaftlern erwarten, den neuen Geist des kollaborativen Arbeitens, das der Digital Turn mit sich bringt, anzueignen. In dieser Wissenschaftspolitik kollidiert die Grundauffassung des geisteswissenschaftlichen Primats der intellektuellen Monade, deren epistemologisches Schaffen in der Introspektion oder der Auseinandersetzung des einzelnen Geistes mit der Theorie und Empirie entsteht (vgl. Reagor et al., 1978) mit dem naturwissenschaftlichen Primat des Erlangens von Wissen aus dem Bezwingen des Faktischen mittels kollektiver Anstrengung (vgl. Knorr-Cetina, 2002, S. 227–235). Ungeachtet der symbolischen Kämpfe, die im wissenschaftlichen Feld<sup>44</sup> zwischen an den Rand gedrängten Geisteswissenschaften und zentral positionierten Naturwissenschaften, die weiterhin vom Ruf unanzweifelbarer Notwendigkeit zehren, toben, ist es der Überlegung wert, welchen Gewinn die Geisteswissenschaften durch die Adaption der Kollaboration als Modell für Wissensgewinn machen können. Dabei stellt sich Kollaboration für die Geisteswissenschaften als neue Grundaktivität wissenschaftlichen Arbeitens dar, die innerhalb des methodologischen Kanons erst etabliert und wie alle tradierten Methoden zuerst erlernt werden muss.<sup>45</sup> Warum Fördereinrichtungen wie die DFG oder das BMBF die Übernahme dieses 'neuen Geistes' von den Geisteswissenschaften fordern, sollte Gegenstand einer längst überfälligen Diskussion sein, die jedoch an anderer Stelle geführt werden muss.

---

44 Zum Feldbegriff vgl. Bourdieu (2005), Bourdieu (2008) sowie Bourdieu (2002) für seine Anwendung auf das soziale Feld der Wissenschaften und Hochschulen.

45 Short (2005, S. 24) vertritt die Auffassung, soll Kollaboration für das geisteswissenschaftliche Paradigma fruchtbar gemacht werden, sei es sinnvoll, dieses Phänomen genauer zu untersuchen.

### 3. Virtuelle Forschungsumgebungen als Entwicklungsfeld

Aus den vorgestellten Sets wissenschaftlicher Primitive, der technisch geförderten neuen Arbeitsformen und der institutionellen Fördersituation bezüglich technisch unterstützter Arbeitsformen leiten sich Implikationen für Virtuelle Forschungsumgebungen ab, die berücksichtigt werden müssen, wenn Bibliotheken sich Virtuelle Forschungsumgebungen als Entwicklungsfeld erschließen wollen.<sup>46</sup> Es ergeben sich drei Komplexe von Anforderungen, die von verschiedenen Seiten an eine Virtuelle Forschungsumgebung herangetragen werden und daher beim Aufbau solcher Infrastrukturen zu berücksichtigen sind. Der erste Komplex bezieht sich auf die direkte Unterstützung wissenschaftlicher Primitive. Um wissenschaftliches Arbeiten tatsächlich sinnvoll unterstützen zu können, müssen Virtuelle Forschungsumgebungen möglichst alle Primitive des jeweiligen (hier also geisteswissenschaftlichen) Wissenschaftsparadigmas abbilden, um die zugehörigen Arbeitsschritte eines Forschungsvorhabens bestmöglich zu fördern. Darüber hinaus müssen sie auch dem institutionellen Rahmen der Wissenschaftsprofession Rechnung tragen, indem sie die daraus abgeleiteten Aufgaben des Wissenschaftsmanagements wie etwa Projektmanagement, Reputationsgenese und Mittelbewirtschaftung sowie institutionell eingeforderte Arbeitsformen wie Kollaboration und die damit in Zusammenhang stehenden sozialen Implikationen unterstützen. Daraus ergibt sich ein institutionell-sozialer Anforderungskomplex, der sich mit den wissenschaftlichen Anforderungen an die digitale Umgebung verzahnt. Diese beiden Komplexe wirken sich direkt auf die technischen Anforderungen, die eine solche Forschungsinfrastruktur erfüllen sollte, aus.<sup>47</sup> Abbildung 2 (S. 31) visualisiert diese drei Komplexe, die von verschiedenen Seiten an eine Virtuelle Forschungsumgebung herangetragen werden und daher beim Aufbau solcher Infrastrukturen zu berücksichtigen sind. Einige dieser Anforderungen durchdringen auch mehrere Komplexe (bspw. gehört Rechtemanagement zu Komplex 2 und 3). Deshalb soll diese Einteilung nicht als starre Zuweisung, sondern mehr als

---

46 Die DINI-Arbeitsgruppe für Virtuelle Forschungsumgebungen, die aus Vertretern von Bibliotheken, Wissenschaft und Rechenzentren besteht, sammelt seit 2009 Anforderungen an solche Umgebungen (vgl. Meyer, 2011, S. 39).

47 Da Virtuelle Forschungsumgebungen nicht unabhängig von ihrer technischen Infrastruktur wie etwa Grids oder Technologien des Semantic Web existieren können, soll der Begriff *Virtuelle Forschungsumgebung* für die kommenden Ausführungen die zugrundeliegende technische Infrastruktur miteinschließen.

Visualisierungshilfe betrachtet werden. Im Folgenden werden diese drei Anforderungskomplexe diskutiert, woraufhin Vorschläge für ihre Unterstützung gemacht werden.

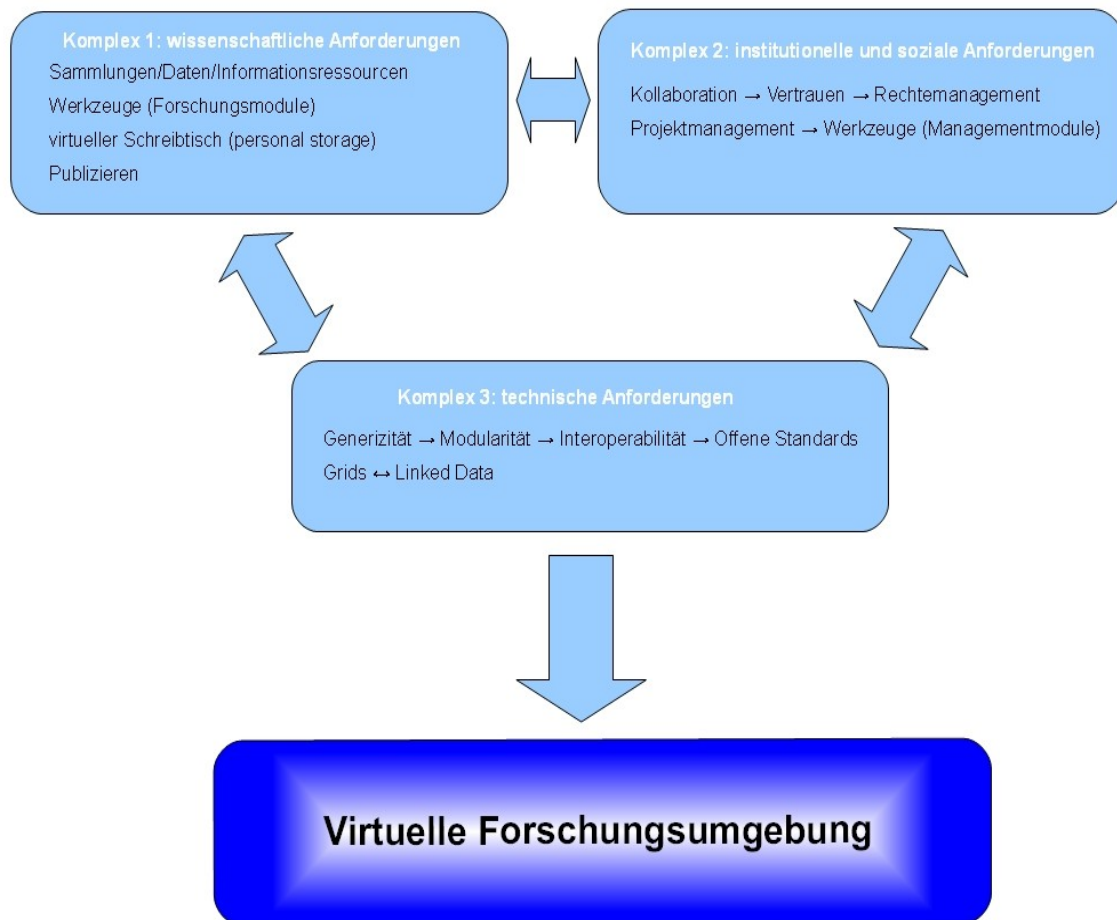


Abbildung 2: Zentrale Anforderungskomplexe an Virtuelle Forschungsumgebungen

### 3.1 Wissenschaftliche und institutionelle Anforderungen

Wissenschaftliche und institutionelle Anforderungen an die Forschungsarbeit beeinflussen sich durch die Überantwortung von Forschungsgeldern und wissenschaftlichem Qualitätsmanagement häufig gegenseitig. Aufgrund des engen Zusammenhangs dieser beiden Anforderungskomplexe, werden sie in ihren Auswirkungen auf die Leistungsanforderungen Virtueller Forschungsumgebungen gemeinsam diskutiert. Aus dieser engen Verknüpfung, die stark mit der

institutionellen Forderung nach Interdisziplinarität in den Forschungsaktivitäten einhergeht, ergibt sich für Virtuelle Forschungsumgebungen die Anforderung, disziplinübergreifend angelegt zu sein. Dieser disziplinübergreifende Ansatz bietet zwei relevante Vorteile. Einerseits können sich aus der interdisziplinären Einsetzbarkeit neue Forschungsmöglichkeiten, Forschungsfragen und sogar ganz neue Forschungsfelder ergeben (z.B. in der Geolinguistik oder Geschichtswissenschaft oder bei der linguistischen Untersuchung von Zeit- und Massenphänomenen wie etwa geographischen Wortverteilungen in physischen und virtuellen Räumen oder sogenannten *Wortkarrieren* (vgl. Liao et al., 2010; Gregory et al., 2011; Aschenbrenner et al., 2007, S. 17; Henn-Memmesheimer, 2006). Andererseits muss dazu nur ein einziges technisches Grundgerüst entwickelt werden, das durch zusätzliche Forschungsmodule, Werkzeuge, die jeweils sehr spezifische Methoden für einzelne Disziplinen zur Verfügung stellen, projekt- und fachspezifisch angepasst werden kann, so dass die wissenschaftlichen Primitive des Paradigmas sowie die speziellen der jeweiligen Disziplin unterstützt werden. Auf diese Weise ergibt sich eine nahezu universelle grundsätzliche Einsetzbarkeit, die trotzdem ein Maximum an Anpassung und Flexibilität für das einzelne Projekt bietet (vgl. American Council of Learned Societies, 2006, S. 8).

Aufbauend auf dieser interdisziplinären Grundlage, erfordert wissenschaftliches Arbeiten im digitalen Raum die Unterstützung und Abbildung der wissenschaftlichen Primitive des betreffenden Wissenschaftsparadigmas in der virtuellen Forschungsumgebung. Zunächst ist es also notwendig, technische Möglichkeiten zu bieten für die Zusammenstellung persönlicher projektspezifischer Korpora und Literatursammlungen, respektive den Zugriff auf bestehende Korpora. Für die Zusammenstellung solcher persönlichen Sammlungen, die das Fundament geisteswissenschaftlicher Forschung bilden, ist es wichtig, Informationsressourcen – seien es Digitalisate und bestehende Korpora, aber auch Datenbanken und sogar Bibliothekskataloge – aus der Virtuellen Forschungsumgebung heraus erreichbar zu machen. Dies ist notwendig für das Entdecken neuer Forschungsfragen, der Methoden zu ihrer Erforschung und das Zusammenstellen persönlicher Sammlungen, auf deren Basis die Anwendung der wissenschaftlichen Primitive Lesen, Annotieren, Wiederlesen und Vergleichen stattfinden kann.



Wichtig für die Verwendung, Analyse und Nachnutzung solcher Korpora ist auch ihr wissenschaftlicher Sammlungskontext. Virtuelle Forschungsumgebungen müssen solchen Sammlungen Rechnung tragen können, indem sie deren Präsentation im relevanten Kontext ermöglichen, z.B. durch die Verknüpfung der Ressourcen mit Quellenkommentaren und Annotationen, die eine Bearbeitung erst ermöglichen. Auch in der Linguistik entstehen beforschbare Primärdaten eigentlich erst dadurch, dass digitalisierte Texte annotiert und durch Informationsanreicherung in Form von Kommentaren in einen Kontext gestellt werden (vgl. Palmer et al., 2009, S. 42). Dazu sind als *Forschungsmodule* zu bezeichnende modular einsetzbare Werkzeuge nötig, die Analysen und Annotationen ermöglichen, die ohne Computerunterstützung nicht denkbar wären.<sup>48</sup> Gleichmaßen notwendig ist die Möglichkeit, sowohl neue Forschungsdaten als auch auf ihnen basierende Publikationen direkt aus der digitalen Umgebung heraus zu veröffentlichen und so das dabei entstandene Wissen in die Wissenschaftsgemeinschaft zu tragen. Damit geht auch die Notwendigkeit eines „virtuellen Schreibtisches“ (*virtueller Desktop*) einher, der es dem Wissenschaftler möglich macht, die eigene Kollektion von Daten, Informationen und Werkzeugen im digitalen Forschungsraum zu speichern, vor unberechtigtem Zugriffen zu sichern und neue Forschungsergebnisse oder Bearbeitungen der jeweiligen Korpora elektronisch zu veröffentlichen. Darüber hinaus soll die Bedienbarkeit der Forschungsumgebung und des virtuellen Schreibtisches so intuitiv sein, dass auch wenig technikaffine Wissenschaftler in diesem digitalen Raum arbeiten können.<sup>49</sup> Entsprechend hoch sind auch die Anforderungen der eHumanities in Bezug auf die Langzeitverfügbarkeit der Daten, ihre dauerhafte Referenzierung, die damit verbundene Qualität der Metadaten und ihre technische Absicherung (vgl. Neuroth et al., 2007, S. 273).<sup>50</sup>

---

48 Eine große Sammlung verschiedener linguistischer Werkzeugprogramme findet sich unter <http://www.clarin.eu/vlo/> und vermittelt einen Überblick über die spezifischen Anforderungen der Linguistik an wissenschaftliches Arbeiten im digitalen Raum. Auch die Korpora des Deutschen Textarchivs und des COSMAS II-Projekts sollen in absehbarer Zeit in diese zentrale Sammlung von Korpora eingehen.

49 Dies erfordert die Entwicklung von *graphical user interfaces* (GUI) mit sehr guter *usability*. Sowohl die Entwicklung solcher GUIs als auch deren Usability-Testing sollte von Bibliotheken bzw. Bibliotheksinstituten aufgrund des direkten Kontakts zur Zielgruppe wissenschaftlich Arbeitender übernommen werden.

50 Die dauerhafte Zugänglichkeit und Identifizierbarkeit geisteswissenschaftlicher Forschungsdaten ist auch eine Kernaufgabe des linguistischen Forschungsnetzwerks CLARIN.

Soll der neue wissenschaftliche Primitiv des Zusammenarbeitens in den Forschungsprozess integriert werden, ist seine Abbildung in Form von technischen Strukturen für Kommunikation und Datenaustausch bzw. dezentralen Zugriff auf verteilte Forschungsdaten innerhalb einer Virtuellen Forschungsumgebung zwingend erforderlich. Zusammenarbeit setzt jedoch auch ein in den Geisteswissenschaften neuartiges Vertrauensverhältnis der Kollaborierenden untereinander voraus, das in den monadisch geprägten Geisteswissenschaften zuerst geschaffen werden muss. Die Bildung dieses Vertrauens kann durch die Implementierung eines differenzierten Rechtemanagement, durch das kontrolliert werden kann, wer Zugriffsberechtigungen auf welche Daten in welchem Projektstadium hat und welche Aktionen – Einsehen, Teilen Kopieren, Bearbeiten, Löschen, Veröffentlichen – an welchen Projekten bzw. Forschungsobjekten durchführen darf, unterstützt werden. Blanke et al. (2011, S. 5) exemplifizieren ein solches vertrauensbildendes – wenn auch immer noch nicht ausreichend differenziertes – Rechtemanagement anhand der Virtuellen Forschungsumgebung TEXTvire<sup>51</sup>:

„TEXTvire also allows for the controlled publication of work-in-progress into restricted repository spaces that are exclusive to individuals or small groups, before the resource is actually made public. As we saw in Section 5, gMan offers similar services, supporting the creation of virtual collections, which provide a mechanism allowing a researcher to share their work, including the relevant research material, annotations and links, with selected colleagues, who in turn add their own annotations and links that may confirm, extend or contest the researcher’s conclusions. A researcher has full control over whether to keep results private, deliver them to a broader group, or indeed make them entirely public. In this way a scholarly dialogue is created and recorded. This could also facilitate new forms of publishing in the humanities, in which readers would have access to the reasoning process that lies behind conclusions, enabling them to validate it – the acceptance of humanities research often depends on provenance of information and on peer assessment – and perhaps criticise it.“

Im Zusammenhang mit der institutionellen Förderstruktur stehen auch Anforderungen des Wissenschafts- und Projektmanagements, das mit Forschung einhergeht. Auch hier ist es sinnvoll, beispielsweise Werkzeuge anzubieten, mit denen Arbeitsschritte jeder Art in Form von Workflows abgebildet werden können, um die Projektentwicklung, aber auch die Mittelbewirtschaftung sowie relevante Termine mittels solcher *Managementmodule* zu visualisieren und auf diese Weise besser

---

51 TEXTvire ist eine Virtuelle Forschungsumgebung für die Geisteswissenschaften in Großbritannien und baut explizit auf dem Konzept von TextGrid auf (vgl. TEXTvire, 2012).

kontrollieren zu können.<sup>52</sup> Die institutionelle Seite wird auch begleitet von der Forderung nach technischer Offenheit der neuentwickelten bzw. neuzuentwickelnden Systeme mit dem Ziel, dass Informationssysteme und Primärdatenrepositorien interagieren können und somit Anschlussfähigkeit für Virtuelle Forschungsumgebungen bieten. Dabei muss die Nachnutzbarkeit von Daten und Ergebnissen durch die Generizität und Interoperabilität der neuen Systeme gewährleistet sein, die durch die Orientierung an offenen Standards und Quelloffenheit erreicht wird (Deutsche Forschungsgemeinschaft, 2011b).

### **3.2 Technische Anforderungen**

Grundsätzlich zeigt sich, dass die einzelnen Disziplinen innerhalb eines Wissenschaftsparadigmas immer ähnliche Anforderungen an eine Virtuelle Forschungsumgebung stellen und die institutionellen Anforderungen des Wissenschaftsmanagements für beide großen Wissenschaftsparadigmen dieselben sind. Statt für jedes Paradigma und jede Einzeldisziplin darin eine oder sogar mehrere Virtuelle Forschungsumgebungen zu entwickeln, die durch mangelnde Anschlussfähigkeit und hohen technischen Unterhaltungsaufwand zu Insellösungen zu werden drohen, vereinfachen generische Umgebungen, in die disziplinspezifische Module und Ressourcen flexibel integriert werden können, den Unterhaltungs- und Entwicklungsaufwand enorm. Aus der Anforderung nach disziplinübergreifender Einsetzbarkeit leitet sich für den technischen Aufbau Virtueller Forschungsumgebungen folglich die Notwendigkeit von Generizität und Modularität ab, die durch eine interoperabel gestaltete Infrastruktur als Basis der gesamten Virtuellen Forschungsumgebung möglich wird.

Interoperabilität leitet sich einerseits aus der in Kapitel 1.1 vorgestellten Natur der Grids als Protokollarchitekturen zum Datenaustausch ab. Nach Blumauer et al. (2006, S. 18) bezeichnet Interoperabilität im Kontext Virtueller Forschungsumgebungen „einen Zustand der Vereinheitlichung und Vereinfachung von Geschäftsprozessen und IT-Architekturen“ mit dem Zweck, „die Interaktion von dispersen Datenbeständen und Anwendungen auf technischer [...] und semantischer

---

<sup>52</sup> Ein Workflow-Tool, das eine solche Funktionalität bietet, ist z.B. Taverna (Taverna, 2012). Einen Überblick zu seinen Funktionalitäten mit Beispielen bieten Wolstencroft et al. (2009).

Ebene zu ermöglichen, ohne dass die Autonomie der einzelnen Teilsysteme aufgehoben wird“. Relevant ist diese Interoperabilität bei allen Szenarien, in denen „interagierende[n] Akteure unterschiedliche Datenformate, Terminologien oder Definitionen verwenden und die Art und Intensität des Datenaustausches mit dem jeweiligen Kontext variiert [...]“ (ebd.). Sowohl die Infrastruktur als auch die darauf aufbauenden digitalen Umgebungen, mit denen der Endanwender direkt in Kontakt steht, sollen durch die Orientierung an Offenen Standards bei der Entwicklung von Diensten, Werkzeugen und Metadaten so gestaltet werden, dass eine Struktur entsteht, die sich auf einfache Weise an neue technische und wissenschaftliche Anforderungen anpassen lässt (vgl. Neuroth et al., 2007). Auf diese Weise kann ein nahtloser Zugang zu Datenrepositorien gewährleistet werden (vgl. American Council of Learned Societies, 2006, S. 8), deren enthaltene Daten dann mittels disziplinspezifischer oder -übergreifender Werkzeuge, wie z.B. Annotationstools, aus den Forschungsmodulen bearbeitet und analysiert werden können. Diese Interoperabilität erfordert sowohl die kontinuierliche Entwicklung anbieterunabhängiger, Offener Standards für Dokumentenmodell und Datendokumentation als auch die damit in Zusammenhang stehende Softwareentwicklung nach dem Open-Source-Prinzip (vgl. American Council of Learned Societies, 2006, S. 37). Diese offenen Strukturen ermöglichen die Entwicklung generischer digitaler Umgebungen, die für die unterschiedlichen Anwendungsszenarien verschiedener Wissenschaftsparadigmen verwendbar sind. Von diesen Strukturen ausgehend, können sowohl unterschiedliche Forschungs- als auch Managementmodule für die Unterstützung vieler verschiedener Forschungsprozesse in die Virtuelle Forschungsumgebung integriert werden. Tabelle 2 (S. 37) fasst die wissenschaftlich-institutionellen und technischen Anforderungen an Virtuelle Forschungsumgebungen zur Übersicht zusammen.

wissenschaftlich-institutionelle Anforderungen	technische Anforderungen
Abbildung wissenschaftlicher Primitive	Generizität
modulare Integration von Informationsressourcen und Forschungswerkzeugen	Modularität
Kommunikationsunterstützung für Zusammenarbeit	Interoperabilität
intuitive Bedienbarkeit	Berücksichtigung Offener Standards
Publikationsunterstützung	
differenziertes Rechtemanagement zur Vertrauensbildung	
Workflow-Unterstützung	
virtueller Desktop	

Tabelle 2: Anforderungen an Virtuelle Forschungsumgebungen

### 3.3 Linked Data als Chance für interpretationsgetriebene Aussagesysteme

Eine besondere Rolle kommt in diesem Kontext der Interoperabilität, Zugänglichkeit und uneingeschränkter Austauschbarkeit der Forschungsdaten selbst zu, die auf der Suche nach neuen Erkenntnissen mittels der Kombination wissenschaftlicher Primitive immer wieder remodelliert und neu miteinander verknüpft werden müssen (vgl. Short, 2005, S. 22). Über Grid-Technologie zum Austausch und Bearbeiten von Forschungsdaten hinaus könnte sich die Verknüpfung verschiedener Korpora und weiterer Informationen zu Linked Data als sehr fruchtbarer Ansatz zur qualitativen Verbesserung der Forschungsdaten und zur Unterstützung des wissenschaftlichen

Erkenntnisprozesses selbst erweisen.<sup>53</sup> Dieses Konzept der semantischen Verknüpfung von Informationsobjekten geht davon aus, dass Daten überall im Internet und WWW (aus Datenbanken, von Webseiten etc.) mittels URI eindeutig referenziert, durch XML, RDF und OWL semantisch angereichert und logisch miteinander verknüpft werden, so dass sie und die Verknüpfungen zwischen ihnen automatisch verarbeitbar werden (vgl. W3C, 2012; Berners-Lee, 2009; Shadbolt et al., 2006) mit dem Ziel, Daten über die Grenzen von Domänen, Organisationen und des World Wide Web hinweg so zu integrieren, dass allein schon durch ihre automatische Abfrage und Verarbeitung neue Informationen zu gewinnen sind (vgl. von Lucke et al., 2010, S. 3).<sup>54</sup> Der Einsatz von Linked Data ist unabhängig von Grids und hat den Vorteil, dass durch die semantische Verknüpfung von Informationen schneller Wissen entstehen kann, indem „fragmentierte, disperse Informationsbestände“ zusammengeführt werden können, um „entscheidungsrelevante Informationen zur Verfügung“ zu stellen (Blumauer et al., 2006, S. 22). Daher ist die Notwendigkeit von Grids in Verbindung mit hochspezialisierten Virtuellen Forschungsumgebungen für geisteswissenschaftliches Arbeiten kritisch zu betrachten, denn die zu verarbeitenden Datenmengen sind einerseits in der Regel nicht so groß, dass es der Rechenkapazität von virtuellen Supercomputern bedarf, um sie zu analysieren. Andererseits bietet die semantische Verknüpfung von Informationsobjekten den effizienteren Weg zu neuem Wissensgewinn, da sie das Auffinden und Interpretieren disperser Datenbestände deutlich erleichtern.

Weiterhin sollte die Generizität der Systeme deutlich über die Basisinfrastruktur hinausgehen. Da viele geisteswissenschaftliche Disziplinen Forschungsmodule benötigen, die auch in anderen Fächern Verwendung finden können, wie etwa GIS-Systeme oder historische Wörterbücher, die sowohl in der Geschichts- als auch Sprachwissenschaft und Geographie genutzt werden, wie z.B. in der

53 Einen tieferen Einblick in das Konzept Linked Data im Zusammenhang mit dem Semantic Web vermitteln (W3C, 2012, W3C, 2010b, W3C, 2010a; Linking Open Data cloud diagram, 2011; O'Reilly, 09/30/2005; Berners-Lee et al., 2001; Shadbolt et al., 2006; Berners-Lee, 2009; Miller, 31.10.2011).

54 Durch ihre semantische Annotationen und logischen Verknüpfungen sind Linked Open Data Teil des Semantic Web nach Berners-Lee et al. (2001). Sie machen Bedeutung für Maschinen insofern erschließbar, dass aus der semantischen Beschreibung der Daten heraus maschinell logische Schlüsse gezogen werden können, die für Menschen wiederum bedeutungshaft interpretierbar sind (ebd.). Dadurch erweitert Linked Data das web 2.0 zum web 3.0 (vgl. Gradmann, 2009).

geschichtswissenschaftlichen Analyse von Ortsnamen in Korpora (vgl. Gregory et al., 2011) sowie in der Sprachkartographie wie beim Digitalen Wenker-Atlas (DiWA, 2012; Wörterbuchnetz, 2011; DWDS, 2012) , ergibt sich kaum die Notwendigkeit, verschiedene, hochspezialisierte Virtuelle Forschungsumgebungen für alle erdenklichen Wissenschaftsdisziplinen zu entwickeln. Gerade über den Ansatz der Generizität<sup>55</sup> könnten Umgebungen geschaffen werden, die Forschungsmodule anbieten, die bis auf geringe Anpassungen von unterschiedlichen geistes- und naturwissenschaftlichen Disziplinen gleichermaßen genutzt werden können. Das hervorstechende und bislang technisch kaum zu bewältigende Spezifikum der Geisteswissenschaften gegenüber den Naturwissenschaften liegt an dieser Stelle eher in der mangelnden Abbildbarkeit ihrer Methoden durch informationstechnische Verfahren: geisteswissenschaftliche Methoden und die darauf basierenden Aussagesysteme sind zumeist stark interpretationsgetrieben.<sup>56</sup> Bislang sind jedoch höchstens Vorstufen von Interpretation für Maschinen abbildbar, weshalb für die Kernmethoden geisteswissenschaftlichen Forschens noch keine automatisierten Verfahren zur Verfügung stehen, die als Forschungsmodule in Virtuelle Forschungsumgebungen implementiert werden könnten.

Verstehen und Interpretieren im menschlich-kognitiven Sinne ist also nach wie vor nicht an Maschinen heranzutragen.<sup>57</sup> Jedoch können Softwareapplikationen komplexe Informationszusammenhänge durch die graphenbasierte Verbindung im Semantic Web für Menschen einfacher bedien- und durchsuchbar gemacht und zusätzliche Kontextinformationen zu Daten angeboten werden, die zumindest die Informationssuche effizienter machen (vgl. Blumauer et al., 2006, S. 22).<sup>58</sup> Das europäische Forschungsnetzwerk CLARIN (CLARIN, 2012) etwa nutzt bereits

---

55 Das hier verwendete Konzept der Generizität lehnt sich an das Konzept der generischen Programmierung an. Ziel ist es, für die Softwareentwicklung Funktionen möglichst allgemein zu entwerfen, um Software-Bibliotheken wiederverwendbar zu gestalten (vgl. Wikipedia Generische Programmierung, 17.04.2012).

56 Prominente Beispiele für interpretative Ansätze sind Konversations- oder Diskursanalyse (vgl. Keim, 2002 und Schneider et al., 2006) sowie die in der verstehenden Soziologie beheimatete dokumentarische Methode (vgl. Bohnsack, 2007).

57 Gradmann (2009, S. 5) postuliert dazu, die moderne Informationswissenschaft müsse sich angesichts der unendlichen Mengen semantischer Einheiten im Netz, die momentan kaum effektiv nutzbar sind, erneut mit dem Verstehenskonzept im digitalen Raum auseinandersetzen.

58 Weitere Beispiele für die Anwendung Linked Open Data im Wissenschaftsbereich sind etwa Clarke et al. (2009) und Engels et al. (2009). Eine erweiterte Sammlung von Use Cases bietet W3C (2010b).

Semantic Web-Technologien<sup>59</sup>, um linguistische Korpora interoperabel zu gestalten und zielt darauf ab, die angeschlossenen Organisationen per Grid zu einer virtuell integrierten Domain zusammenzuschließen. CLARIN möchte seine Forschungsressourcen, Dienste und Werkzeuge über das Internet zugänglich machen und seiner Nutzercommunity verschiedene Zugangswege zu und Trainingsmöglichkeiten mit den linguistischen Daten zur Verfügung stellen und gleichzeitig die Möglichkeit bieten, durch seine offene Infrastruktur weitere Informations- und Forschungsressourcen sowie zusätzliche Dienste zu integrieren (ebd.). Durch die Verbindung vieler Ressourcen als Linked Open Data können wesentlich größere Datenmengen für die Wissenschaft hochwertigere Informationen in kürzerer Zeit bieten und die Informationsflut im Netz besser beherrschbar machen, indem die bislang bestehenden, von einander getrennten „Datensilos“ durch Verknüpfung untereinander aufgelöst und Informationen aus ihnen schneller extrahiert werden kann, um mehr Wissen zu generieren (vgl. Berners-Lee et al., 2001, S. 37). Dies bietet sich vor allem für die textbezogenen Geistes- und Gesellschaftswissenschaften mit ihren interpretationsgetriebenen Methoden an, deren Korpora durch semantische Annotation gut erfassbar sind<sup>60</sup>.

Betrachtet man die Vorteile, die Linked Open Data für die Wissenschaft mit sich brächte, drängt sich unweigerlich die Frage auf, warum sich das Konzept noch nicht durchgesetzt hat.<sup>61</sup> Es liegen sowohl technische als auch soziale Probleme vor. Die allgemein technischen Umsetzungsprobleme für das Semantic Web und die damit zusammenhängenden Konzepte lassen sich nach Weller (2009, S. 8–10) wie folgt zusammenfassen: Ontologien, die eine semantische Verknüpfung von Daten ermöglichen, sind bisher nur für eng umrissene Themenbereiche entwickelt, da der Aufwand der Erstellung eines stark ausdifferenzierten Wissensrepräsentationssystems sehr hoch ist – für eine universelle Repräsentation des Weltwissens sogar zu hoch. Die bestehenden Einzelontologien sind bislang

---

59 Zur Unterscheidung zwischen Semantic Web und semantischen Technologien s. Blumauer et al. (2006).

60 Auch die Zeitschriftendatenbank stellt ihre Daten jetzt als Linked Data zur Verfügung s. Zeitschriftendatenbank (2012).

61 Diese Frage kann aufgrund des Umfangs und Ansatzes der vorliegenden Arbeit nur in Grundzügen angesprochen werden und muss einer eingehenden Diskussion an anderer Stelle harren. Dennoch soll hier der Hinweis auf basale Probleme bei der institutionellen und technischen Durchsetzung des Konzepts hingewiesen werden.



kaum verknüpfbar. Die Handhabung der immensen Datenmengen, die durch Wissensmodelle entstehen, die in formalen Ontologiesprachen verfasst sind, gestaltet sich noch schwierig. Derzeit bieten Ontologien, deren Einsatz für das Semantic Web überhaupt zur Debatte stehen, kaum mehr Ausdrucksfähigkeit als Thesauri.<sup>62</sup> Darüber hinaus mangelt es nach wie vor an der sogenannten *Killer App*, die Web-Nutzer und Anbieter vom Semantic Web überzeugt. Gleichzeitig wäre die semantische Anreicherung des gesamten WWW eine schier endlose Aufgabe und Standards für Indexierung von Web-Inhalten mit Ontologien sind noch nicht entwickelt. Darüber hinaus scheitert die Durchsetzung der Veröffentlichung von Daten derzeit vor allem an der technischen Unlösbarkeit eines sozialen Problems. Der Umgang mit Primärforschungsdaten erfordert aus wissenschaftlicher Sicht ein hohes Maß an technisch umzusetzender Diskretion. Unabhängig vom Wissenschaftsparadigma des jeweiligen Forschers ist es Wissenschaftlern wichtig, der Öffentlichkeit und anderen Wissenschaftlern keinen oder nur sehr geringen Einblick in Projekte zu geben, die sich noch im Entwicklungsstadium befinden, um den wissenschaftlichen Impact der eigenen Arbeit nicht zu gefährden. Es ist daher nicht erwünscht, dass Außenstehende sehen, woran der Wissenschaftler mit welchen Methoden und vorläufigen Ergebnissen arbeitet. Gerade vor der Fertigstellung bzw. Veröffentlichung von Forschungsergebnissen ist dieses Streben nach maximaler Diskretion stark ausgeprägt und vor dem Hintergrund der Struktur des wissenschaftlichen Feldes nachvollziehbar.<sup>63</sup> Ein derart differenziertes Autorisierungsmanagement, das nur ganz bestimmten Nutzern bzw. Institutionen im WWW Zugriff auf die Linked Open Data einer anderen Institution oder Person gewähren kann, ist im Rahmen der Schlüsseltechnologie des Semantic Web, RDF, noch nicht zufriedenstellend umsetzbar. Andererseits scheitert das Konzept der Linked Open Data – ähnlich den Schwierigkeiten, denen sich das Open-Access-Konzept gegenüber sieht – im wissenschaftlichen Feld an den Positionierungskämpfen zwischen etablierten Wissenschaftlern und wissenschaftlichem Nachwuchs. Während es für etablierte Forscher wichtig ist, ihre

---

62 Es existieren jedoch Ansätze zur Übertragung von Thesauri ins Semantic Web (u.a. Eckert et al., 2008; Eckert et al., 2009; Eckert, 2011).

63 Wissen im Sinne von relevanten oder sogar bahnbrechenden Forschungsergebnissen stellt in diesem Feld durch den Zugewinn an wissenschaftlicher Reputation, also Prestige, die primäre Ressource sozialen Kapitals und somit auch ökonomischen Kapitals dar (vgl. Bourdieu, 2002).

aufgebaute Reputation zu erhöhen, geht es bei Nachwuchswissenschaftlern um den Aufbau des wissenschaftlichen Rufes. Reputation wird in der Regel generiert durch Vortragstätigkeiten auf renommierten Konferenzen und Veröffentlichungen in stark rezipierten Zeitschriften, Reihen und Konferenzbänden. Das kulturelle und soziale Kapital 'Reputation' ist in diesem Szenario einfach zu vermehren, doch aufgrund der Eintrittshürden ins wissenschaftliche Feld schwer aufzubauen. Indem kapitalstarke Wissenschaftler ihre Primärforschungsdaten unter Verschluss halten verhindern sie nicht nur deren parallele Bearbeitung durch andere (kapitalschwächere) Wissenschaftler oder eine kritische Betrachtung der darauf basierenden Forschungsergebnisse, sondern auch den Reputationsaufbau ihrer Feldkonkurrenten durch Ressourcenverknappung. Diese Restriktionen durch Verknappung tragen sich auch in den Publikationsbereich weiter, wo die Eintrittsschwellen zur Veröffentlichung sich für Nachwuchswissenschaftler sehr hoch gestalten.<sup>64</sup>

#### **4. Das „Deutsche Textarchiv“ – vom digitalen Korpus zur Virtuellen Forschungsumgebung**

Nachdem die Voraussetzungen und Anforderungen an Virtuelle Forschungsumgebungen dargestellt wurden, sollen deren Umsetzungsmöglichkeiten am digitalen Korpus des „Deutschen Textarchivs“ (DTA, 2011), das auch verschiedene Forschungswerkzeuge bietet, betrachtet werden. Im Fokus der kritischen Darstellung steht die Abbildung wissenschaftlicher Primitive innerhalb der digitalen Arbeitsumgebung des Deutschen Textarchivs, der angebotenen Werkzeuge zur Bearbeitung der Informationsressourcen sowie Vorschläge und Planungen für die zukünftige Weiterentwicklung der Umgebung.

---

<sup>64</sup> Aufgrund der Komplexität und Relevanz dieses Themas ist seine weitere Diskussion im Rahmen dieser Arbeit nicht möglich. M. E. ist eine eingehende Analyse dieser Positionierungskämpfe für den deutschen Wissenschaftsmarkt auf der Basis Bourdieuscher Theorie dringend notwendig, wenn eine Veränderung des Umgangs mit Linked Data und Open Access in diesem Bereich angestrebt werden soll.

## 4.1 Korpusbeschreibung

Das Deutsche Textarchiv wird seit 2007 von der Berlin-Brandenburgischen Akademie der Wissenschaften (BBAW) und verschiedenen vor allem bibliothekarischen Projektpartnern entwickelt mit dem Ziel, ein disziplinübergreifendes, linguistisch annotiertes und repräsentatives Volltextkorpus deutschsprachiger Texte aus dem Zeitraum von 1650 bis 1900 öffentlich im Internet und als weiterverarbeitbare HTML- und XML TEI P5-Dateien<sup>65</sup> zum Download zur Verfügung zu stellen.<sup>66</sup> Das Projekt wird durch die Deutsche Forschungsgemeinschaft gefördert. Nach der Zusammenstellung einer Titelliste<sup>67</sup> durch das BBAW wurde die Digitalisierung der historischen Erstausgaben der ausgewählten Texte durch verschiedene kooperierende Bibliotheken, die im Besitz der betreffenden Werke sind, vorgenommen.<sup>68</sup> Das derzeit 532 von 650 angestrebten Büchern umfassende Korpus setzt sich jeweils aus den digitalisierten Faksimiles der Vorlagen, den qualitätsgeprüften und manuell korrigierten Volltexten und Anmerkungen dazu sowie den zugehörigen Strukturdaten (bestehend aus Kapitel- und Unterkapitelüberschriften, Absätzen, typographischen Hervorhebungen, Bogensignaturen, Kolummentiteln usw.) zusammen. Texte und Bilddigitalisate wurden wortweise miteinander verknüpft. Alle Volltexte liegen im XML-basierten TEI P5-Format vor und sind mit computerlinguistischen Methoden erschlossen. Die vertiefte linguistische Erschließung erlaubt die Abfrage verschiedener historischer Orthographievarianten eines Wortes über die heute gebräuchliche Wortform. Dabei wurden alle verschiedenen orthographischen Varianten eines Wortes anhand ihrer graphematischen Ähnlichkeiten durch ein eigens dafür entwickeltes Programm ermittelt und unter der aktuellen orthographischen Form zu einem Metalemma

---

65 TEI steht für *text encoding initiative* und ist ein semantic-web-fähiger Standard zur Annotation von Text (dazu u.a. Jannidis, 2009; Zielinski et al., 2009; Schreibman, 2009 und weiterführend Hockey, 1993).

66 Die BBAW verweist darauf, dass der momentane Entwicklungsstand noch als Betaphase zu bezeichnen ist (mündliche Mitteilung vom 04.05.2012).

67 Zu den Auswahlkriterien der infragekommenden Texte und Digitalisierungsvorlagen vgl. <http://www.deutschestextarchiv.de/project/>.

68 Kooperierende Bibliotheken, die ihre Bestände zur Verfügung stellen und die Digitalisierung vornehmen, sind die Staatsbibliothek zu Berlin Preußischer Kulturbesitz, die Zentral- und Landesbibliothek Berlin, die Bibliothek des BBAW und der Arbeitsstelle Deutsches Wörterbuch (ebenfalls BBAW), die Niedersächsische Staats- und Universitätsbibliothek Göttingen zusammen mit dem Göttinger Digitalisierungszentrum, die Herzog August Bibliothek Wolfenbüttel sowie das Max-Planck-Institut für europäische Rechtsgeschichte (vgl. <http://www.deutschestextarchiv.de/project/coop/>).

vereinigt. Um die komplexe Suche nach flektierten Wortformen, semantischen Konzepten, Phrasen und weiteren sprachlichen Mustern zu ermöglichen, wurden alle Begriffe in den Texten tokenisiert, lemmatisiert<sup>69</sup> und anschließend in einer linguistischen Suchmaschine DDC indexiert.

## **4.2 Forschungswerkzeuge**

Das Korpus bietet zwei verschiedene Sucheinstiege, eine linguistische Suchmaschine und einen Viewer für die Arbeit mit den Texten. Weiterhin beinhaltet diese Viewer verschiedene Ansichtsoptionen sowie die Möglichkeit zur Vergrößerung, Tiefenreferenzierung und dauerhaften Annotierung einzelner Textpassagen innerhalb eines Werks. Die genannten Werkzeuge werden im folgenden dargestellt.

### **4.2.1 Sucheinstiege**

Das Deutsche Textarchiv bietet dem Nutzer von der Startseite (Abb. 3, S. 45) ausgehend zwei grundlegende Sucheinstiege in das Korpus: die thematische oder begriffsbezogene Suchanfrage über den Suchschlitz und den Einstieg über ein einzelnes Werk, das aus der Liste der aktuell zugreifbaren Bücher ausgewählt wird.

---

<sup>69</sup> Bei der Tokenisierung werden alle im Text vorkommenden Wortformen aufgelistet (z.B. *fliegst*, *flöge*, *flog*, *geflogen*, *flieg*) und durch einen Lemmatisierer auf ihre jeweilige lexikalische Grundform (hier: *fliegen*) zurückgeführt.

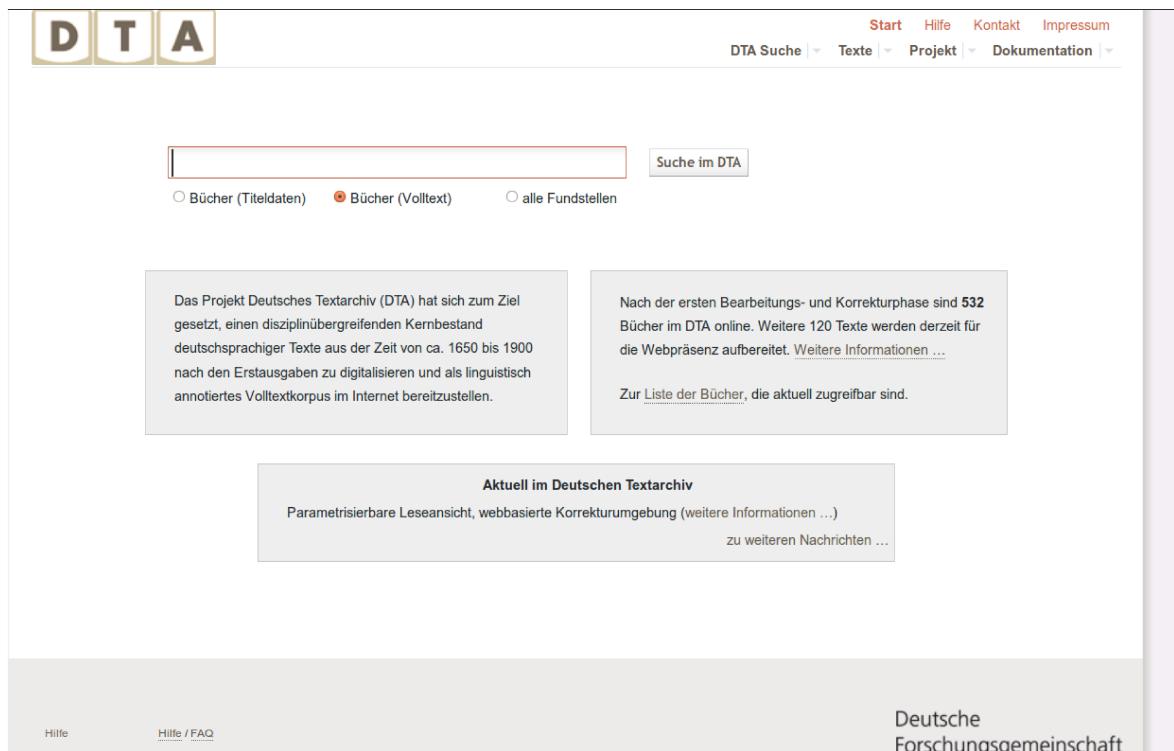


Abbildung 3: Startseite des Deutschen Textarchivs

### Bücherliste

Diese Liste digitalisierter Werke (vgl. Abb. 4, S. 46) bietet dem Nutzer die Möglichkeit, sich einen ersten Überblick über das Korpus zu verschaffen, da es nach Autoren, Titeln und Erscheinungsjahr sortierbar ist. Diese Bücherliste erlaubt einen erweiterten Einblick ins Korpus, indem sie auch eine Zeitleiste der Werke und eine Übersicht über den Digitalisierungsfortschritt im Korpus bietet. Die Zeitleiste, die über die gleichnamige Schaltfläche oberhalb der Bücherliste zugänglich ist, vereinfacht den Überblick über die Autoren, die ins DTA aufgenommen sind und die repräsentierten Epochen. Dazu sind die Geburts- und Sterbedaten der Autoren sowie die Publikationsdaten der Werke des Korpus verzeichnet. Die Zeitleiste vermittelt dem Anwender eine Vorstellung der Epochen, in denen er sich bewegt und der literarischen Entwicklung dieser Zeit. Über die fortschreitende Entwicklung des Korpus selbst informiert der Link „Fortschritt“ oberhalb der Bücherliste. Die dahinterliegende Übersicht zeigt, wie weit Titel, die sich gerade in Bearbeitung für das DTA befinden, bereits eingescannt sind (vgl. Abb. 5, S. 47). Leider ist die Darstellung der Zeitleiste sehr unübersichtlich, so dass eine Orientierung schwerfällt.

Anmelden

**DTA** [Start](#) [Hilfe](#) [Kontakt](#) [Impressum](#)

DTA Suche | [Texte](#) | [Projekt](#) | [Dokumentation](#) |

[Titel \(online\)](#) [Zeitleiste](#) [Fortschritt](#)

Suche im DTA [Suchhistorie](#)

### DTA Bücher Verzeichnis

In der ersten Projektphase des Deutschen Textarchivs werden ca. 650 Werke, die eine repräsentative und disziplinübergreifende Auswahl deutschsprachiger Texte darstellen, in digitaler Form zur Verfügung gestellt. Derzeit können Sie auf die nachstehenden Werke im Deutschen Textarchiv zugreifen.

Mehr Information zum Auswahlverfahren...

Noch in der Digitalisierung befindliche Bände von 1780-1900...

Sortieren Sie die Titelliste per Klick auf Titel, Autor oder Jahr nach Ihren Vorstellungen.

	Autor	Titel	Jahr	Anzeige
1	Lorenz Oken	Abriß des Systems der Biologie	1805	<a href="#">B</a>   <a href="#">T</a>
2	Friedrich Theodor Vischer	Aesthetik oder Wissenschaft des Schönen 1	1846	<a href="#">B</a>   <a href="#">T</a>
3	Friedrich Theodor Vischer	Aesthetik oder Wissenschaft des Schönen 2	1847	<a href="#">B</a>   <a href="#">T</a>
4	Friedrich Theodor Vischer	Aesthetik oder Wissenschaft des Schönen 3	1848	<a href="#">B</a>   <a href="#">T</a>
5	Friedrich Theodor Vischer	Aesthetik oder Wissenschaft des Schönen 4	1851	<a href="#">B</a>   <a href="#">T</a>
6	Friedrich Theodor Vischer	Aesthetik oder Wissenschaft des Schönen 5	1852	<a href="#">B</a>   <a href="#">T</a>
7	Friedrich Theodor Vischer	Aesthetik oder Wissenschaft des Schönen 6	1853	<a href="#">B</a>   <a href="#">T</a>
8	Friedrich Theodor Vischer	Aesthetik oder Wissenschaft des Schönen 7	1854	<a href="#">B</a>   <a href="#">T</a>
9	Friedrich Theodor Vischer	Aesthetik oder Wissenschaft des Schönen 8	1857	<a href="#">B</a>   <a href="#">T</a>
10	Friedrich Theodor Vischer	Aesthetik oder Wissenschaft des Schönen 9	1857	<a href="#">B</a>   <a href="#">T</a>
11	Stefan George	Algabal	1892	<a href="#">B</a>   <a href="#">T</a>
12	Karl Friedrich Gauß	Allgemeine Lehrsätze in Beziehung auf die im verkehrten Verhältnis des Quadrats der Entfernung wirkenden Anziehungs- und Abstoßungskräfte	1840	<a href="#">B</a>   <a href="#">T</a>

Abbildung 4: Liste der zugreifbaren Bücher des DTA

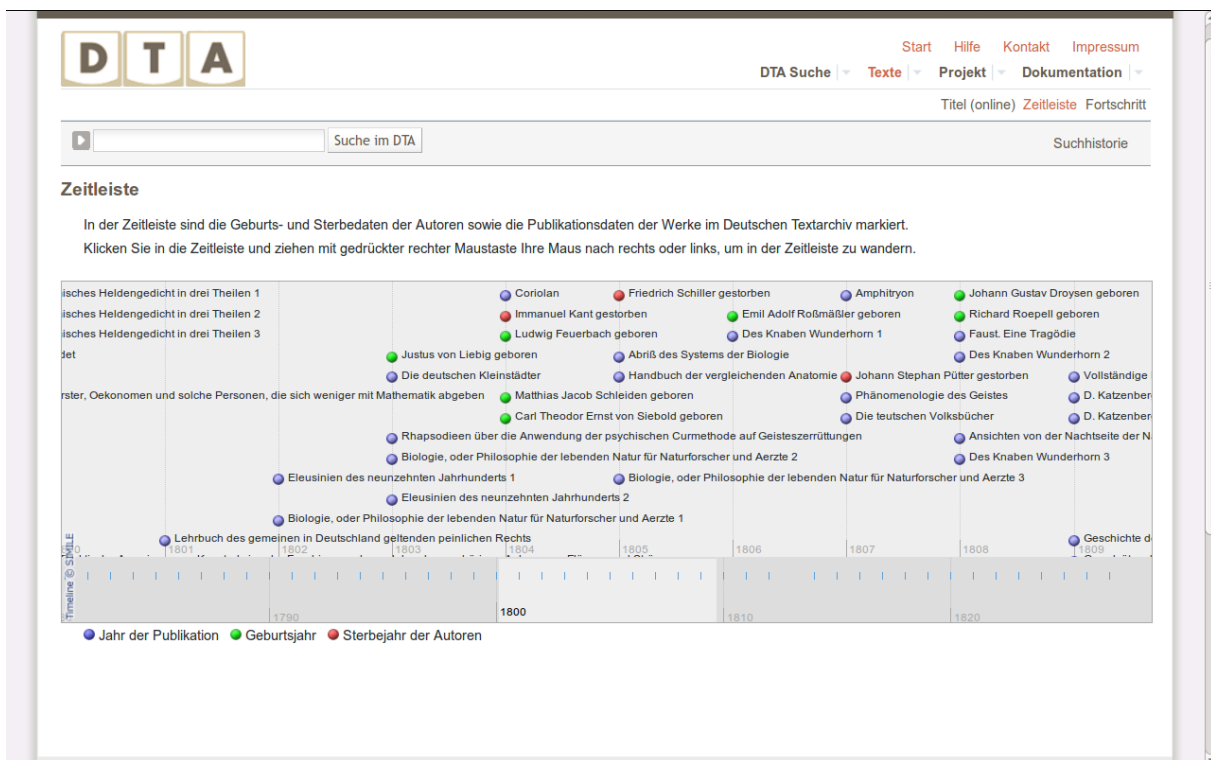


Abbildung 5: Zeitleiste

Durch eine Weiterleitung von der Bücherliste auf den DTA-Viewer ist die direkte Anzeige des jeweiligen Titels als Bilddatei oder Volltext möglich. Der Viewer ist nicht nur das Werkzeug zum Betrachten einzelner Titel des Korpus, sondern bietet dem Anwender zugleich einen virtuellen Schreibtisch. Hier findet der Nutzer verschiedene Ansichtsoptionen für das ausgewählte Werk: den Imagescan, den digitalen Volltext, die XML-Struktur, Kombinationen dieser Ansichtsmodi und eine Überblicksdarstellung aller Bilder des ausgewählten Werkes. Über diesen Einstieg kann sich der Nutzer verschiedene Werke, eine Auswahl von Werken bestimmter Autoren oder Werke derselben Textgattung, eines spezifischen Zeitabschnitts oder Genres zu einem individuellen Korpus zusammenstellen. Bereits über diesen Sucheinstieg wird das entdeckende Suchen und Sammeln (vgl. Blanke et al., 2011) für die Auswahl von passender Literatur für ein Forschungsprojekt unterstützt. Der direkte Zugang zum ausgewählten Werk über den DTA-Viewer bietet einen Überblick über die Metadaten des Titels inklusive seiner Provenienz, URN zur Referenzierung, seine Schrifttype sowie eine Verknüpfung zur Personennamendatei (PND), die den Autor eindeutig identifiziert. Außerdem wird per Link der direkte Download des

Werkvolltextes im HTML- oder XML-Format angeboten sowie der Zugang zu den Bilddateien des Buches, zur Volltext- und XML-Anzeige sowie zur Suche im Buch (vgl. Abb. 6, S. 48). Darüber hinaus stehen eine Ausschneidefunktion, eine Lupenfunktion zur Ausschnittvergrößerung, eine Funktion für die optimale Auflösung und ein Vollbildmodus zur Verfügung.

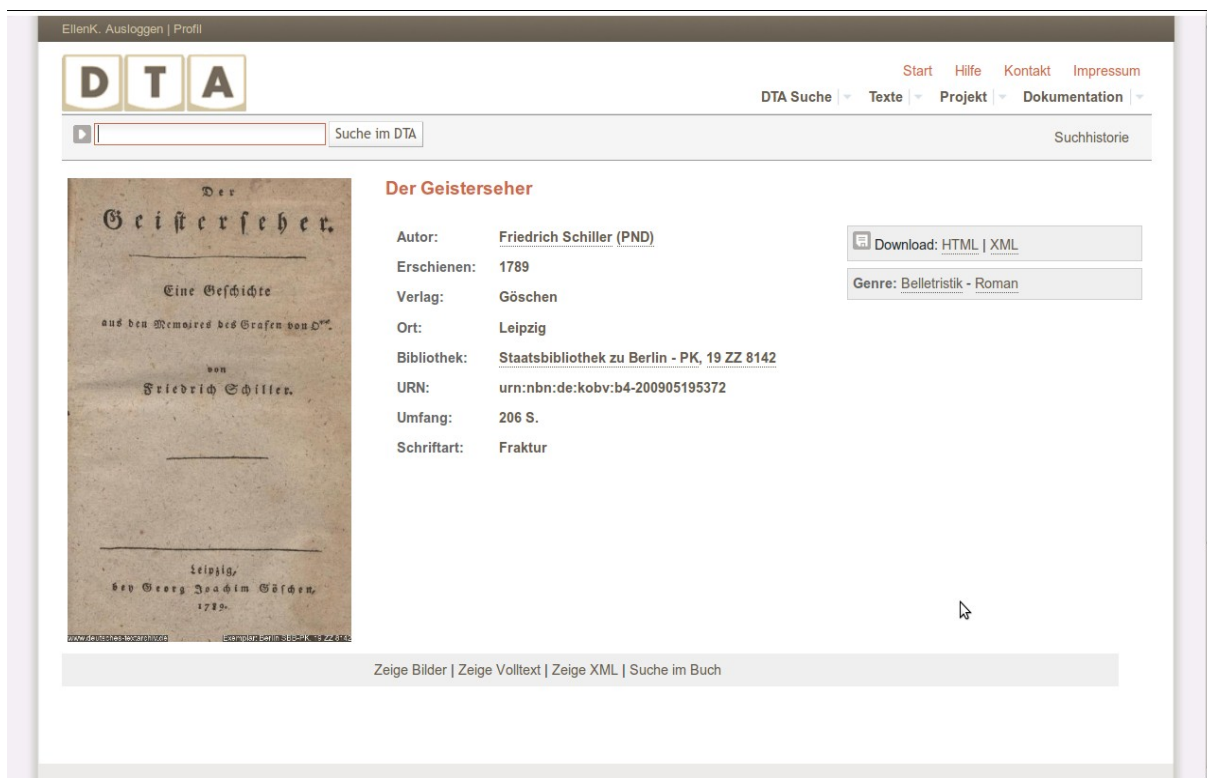


Abbildung 6: Ansicht des Einzeltitels „Der Geisterseher“ von Friedrich Schiller im DTA-Viewer

Alle Texte sind einem Genre und einer Textgattung zugeordnet. Von dieser Einordnung ausgehend kann auch nach weiteren Texten des betreffenden Genres oder derselben Textgattung weitergesucht werden. Nach welchen Kriterien Genre und Gattung vergeben werden, geht aus der veröffentlichten Dokumentation des Archivs jedoch nicht hervor. Da über diese Einordnung bereits ein Forschungsergebnis dieses Korpusprojekts darstellt, sollten die Denk- und Entscheidungsprozesse, die zu den Ergebnissen führten, unbedingt transparent gemacht werden. Mit diesem Ansichtswerkzeug wird nicht nur der Prozess des Entdeckens ähnlicher Titel, sondern auch des Vergleichens von Werken unterstützt.



Der Sucheinstieg über ein bestimmtes Werk kann so zum Finden weiterer ähnlicher oder vergleichbarer Werke verwendet führen.

### Einfache und komplexe Suchanfrage

Wählt der Nutzer den Sucheinstieg über den Suchschlitz der Startseite, kann er sowohl einfache als auch komplexe Anfragen mit der DDC-Suchmaschinensyntax eingeben. Dabei kann die Trefferliste auf die Titeldaten der Werke, die Werke als Volltexte und „alle Fundstellen“, d.h. die konkreten Treffer auf allen Seiten der Volltexte, beschränkt werden. Abbildung 7 (S. 49) zeigt die Beispielabfrage für den einfachen Suchbegriff *Geisterseher*. Es gibt zu dieser Anfrage keine Treffer in den Metadaten („Titeldaten“) der Werke, 22 Volltexte („Bücher“), in denen der Suchbegriff vorkommt, und 27 Fundstellen insgesamt. Sollen bestimmte Typen von Werken gesucht werden, z.B. Handbücher, können diese durch Eingabe des entsprechenden Suchbegriffs und Auswahl des Kartereiters „Bücher (Titeldaten)“ gefunden werden. Bei der Suche in den Titeldaten der Werke wird die linguistische Suchsyntax jedoch nicht unterstützt.

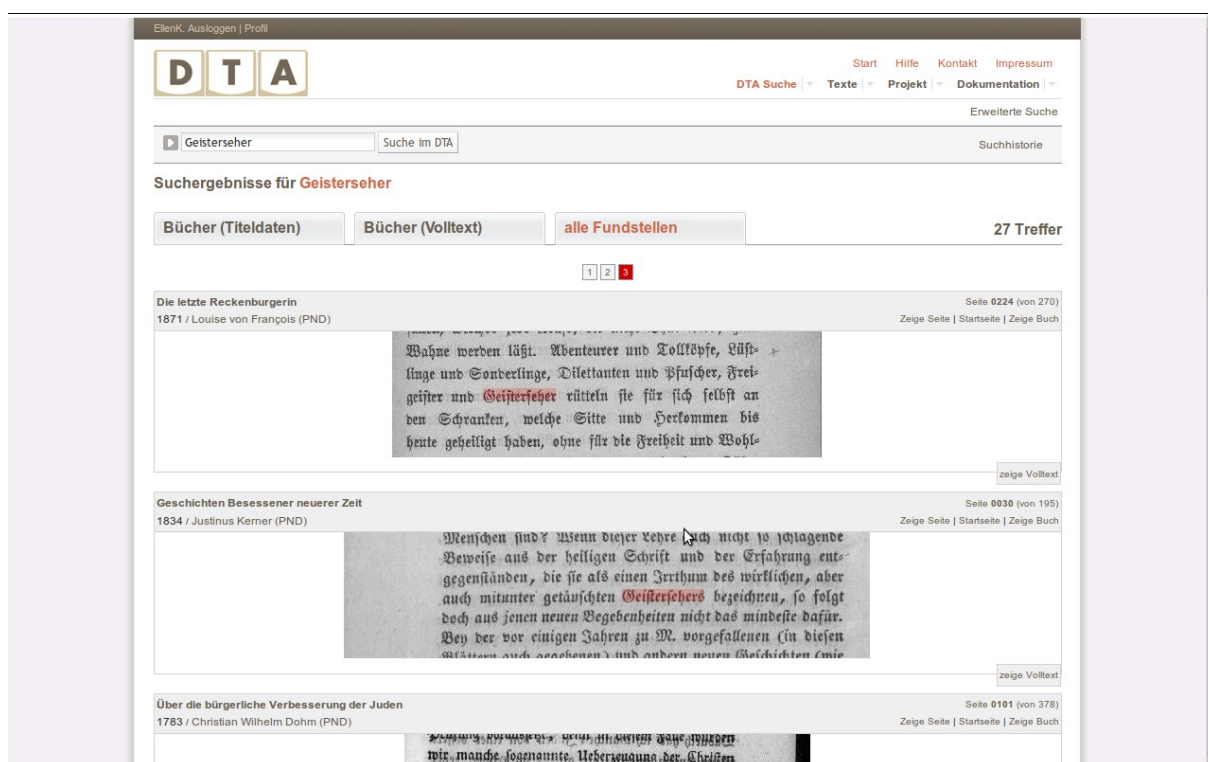


Abbildung 7: Einfache Suchanfrage nach *Geisterseher* mit Anzeige aller Fundstellen

Die einzelnen Treffer können als Volltext oder Einzelseite ausgewählt und betrachtet werden. Die Trefferbegriffe oder -phrasen aus der Suchanfrage sind orange-rot markiert. Über die Ausschnitt-Funktion des DTA-Viewers (Scherensymbol am rechten oberen Bildrand) kann der Nutzer Textausschnitte aus der jeweiligen Seite erstellen und sie für sein persönliches Korpus kategorisieren, annotieren, speichern und mit dauerhaften Lesezeichen versehen (vgl. Abb. 8, S. 50), wofür der Anwender jedoch zuvor ein Nutzerkonto angelegen und sich anmelden muss. Die wissenschaftlichen Primitive Sammeln, Annotieren und Vergleichen werden also in dieser Funktion unterstützt, indem ein selbstangereichertes Korpus angelegt werden kann.

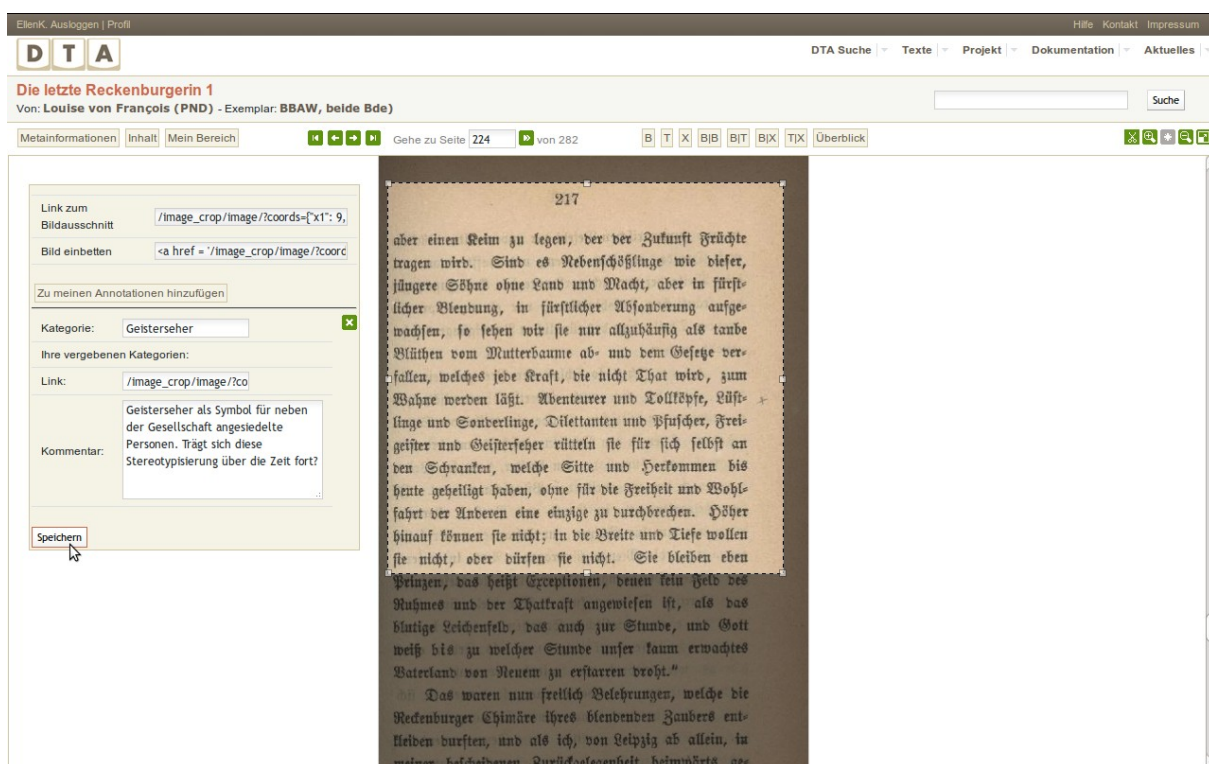


Abbildung 8: Textauswahl mit individueller Kategorisierung und Annotation

Der zeitliche Kontext, das Genre und die Gattung des Textes können über die Schaltfläche „Metainformationen“ eingeblendet werden. Zur Navigation innerhalb des Buches sind über die Schaltfläche „Inhalt“ Strukturdaten anzeigbar, die mit einzelnen Textabschnitten verlinkt sind. Aus der Einzelbildansicht ist auch das Vor- und

Zurücknavigieren im Buch möglich. Abbildung 9 (S. 51) zeigt einen Gesamtüberblick aller Funktionen des DTA-Viewers.

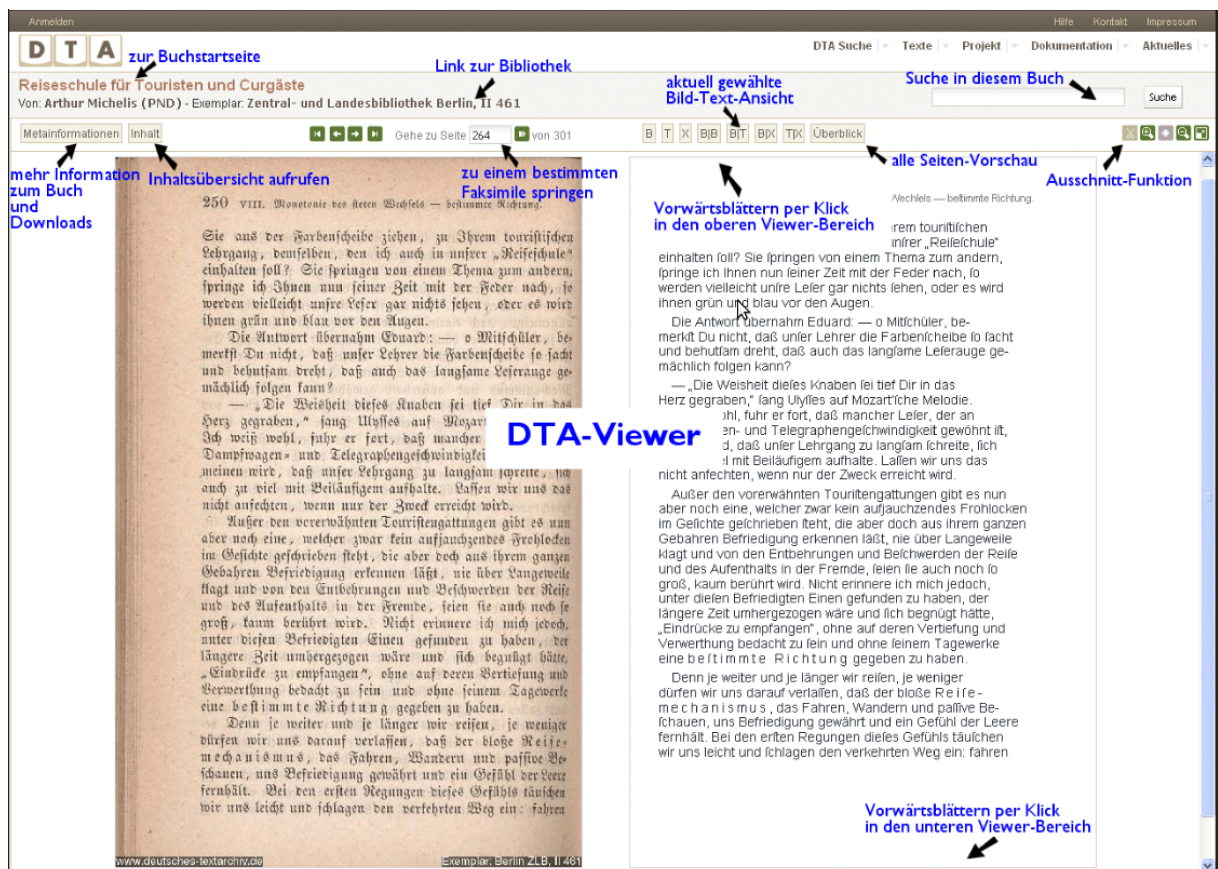


Abbildung 9: Funktionen des DTA-Viewers

#### 4.2.2 Linguistische Suchmaschine

Die linguistische Suchmaschine Dialing/DWDS-Condordancer (DDC) bildet das Herzstück der Werkzeuge, die das DTA anbietet. Sie wurde durch das BBAW ursprünglich für das Digitale Wörterbuch der deutschen Sprache des 20. Jahrhunderts (DWDS, 2012) entwickelt und in das DTA integriert (DTA, 2011). Die Suchmaschine bietet neben der klassischen Volltextsuche auch die Recherche in den Metadaten und eine linguistische Suche über Boolesche Operatoren (UND,

ODER, NICHT) und eine Stemmasuche<sup>70.71</sup> Zusätzlich indiziert die Suchmaschine DDC auch die Metadaten der Werke, was Suchanfragen auf diese erlaubt. Die komplexe Syntax für die Volltextsuche ermittelt zu jeder Eingabe automatisch sowohl andere Wortformen als auch historische Orthographievarianten, doch auch die Suche nach einer bestimmten Wortform (z.B. nur *Mond*, statt *Monde*, *Monden*, *Mondschein*) ist durch Voranstellen eines Kommandozeichens (hier das @-Symbol) möglich. Als linguistische Suchmaschine ist die DDC case-sensitive und unterstützt Rechts- wie Linkstrunkierung sowie Phrasensuche (d.h. die Suche nach einer bestimmten Wortfolge) und Abstandssuche (d.h. andere Wörter zwischen den Begriffen der gesuchten Phrase werden geduldet).<sup>72</sup> Über die Schaltfläche „Erweiterte Suche“ kann der Suchraum auf Metadateneinträge (z.B. Titel, Autor, Jahr, Zeitraum), aber auch auf inhaltliche Kontextfilter (etwa Fußnoten, Marginalien, Bühnenanweisungen etc.) eingeschränkt werden. Außerdem ist die Suche über typographische Kontextfilter, d.h. Drucktypen, möglich. Dabei bezieht sich die Suche in der Regel auf Passagen, in denen entweder die Frakturtype wechselt oder ein Wechsel von Fraktur- auf Antiquaschrift stattfindet. Die Metadaten- und Kontextfilter sind miteinander kombinierbar. Die Ergebnisliste wird in den drei bereits vorgestellten Kategorien „Bücher (Titeldaten)“, „Bücher (Volltext)“ und „alle Fundstellen“ nach Relevanz sortiert ausgegeben. Um die bislang getätigten Suchanfragen zu überblicken, kann die bisherige Suchhistorie über die zugehörige Schaltfläche im oberen Drittel der Bildschirmseite aufgerufen und einzelne Suchanfragen daraus per Klick erneut durchgeführt werden. Um die Suchhistorie, ebenso wie die Annotationsfunktion nutzen zu können, muss der Anwender über sein Nutzerkonto angemeldet sein.

---

70 Die Suche nach Wortstämmen ermöglicht das Auffinden aller Formen eines Verbs ungeachtet der jeweiligen Flexionsform. Dies wird durch die Zusammenführung jeweils aller stammzugehörigen Wortformen zu einem Metalemma im Suchmaschinenindex erzielt. Jeder Indexeintrag wird dazu mit seiner Wortart und dem zugehörigen Wortstamm annotiert.

71 Ähnliche und weitere Möglichkeiten zur differenzierten Such in Volltextkorpora bietet auch COSMAS II.

72 Einen genauen Überblick über die Abfragesyntax mit verschiedenen Beispielen bietet [http://www.deutschestextarchiv.de/help/#part\\_3](http://www.deutschestextarchiv.de/help/#part_3).

## **4.3 Geplante Weiterentwicklung**

### **4.3.1 Aktives Archiv**

Das Deutsche Textarchiv ist in seiner Entwicklung noch im Anfangsstadium begriffen, weshalb noch verschiedene Erweiterungen und Verbesserungen angestrebt werden. Zum einen ist das Deutsche Textarchiv als „aktives Archiv“ (vgl. DTA, 2011) geplant, in dem die Nutzer nicht nur forschen, sondern das sie selbst auch bearbeiten, indem sie sich ihr eigenes Korpus aus den Texten zusammenstellen und Textausschnitte mit dauerhaften Lesezeichen und Annotationen versehen. Dieser Teil des „aktiven Archivs“ ist bereits zum großen Teil umgesetzt. Zur Erweiterung des Archivs steht zur Diskussion, registrierten Nutzern durch das Erweiterungskorpus des Deutschen Textarchivs DTA-E (DTA-E, 2012) die Möglichkeit zu geben, qualitätsgeprüfte Texte nach Absprache mit der BBAW selbst in das Deutsche Textarchiv zu integrieren. Zusätzlich ist ein Erweiterungskorpus für den Zeitraum nach 1900 geplant, das auch Raum für nicht-buchbasierte Texte bietet. Derzeit beinhaltet das DTA-E bereits Leichenpredigten und sogenannte „Christliche Ratgeber“, doch auch soziolinguistische Sprach- oder Textkorpora wären denkbar.

### **4.3.2 Parametrisierbare Leseansicht**

Auch der DTA-Viewer soll um eine Leseansicht für Texte erweitert werden, die das Lesen der Texte am Bildschirm komfortabler gestalten und ein seitenweises Blättern im Text überflüssig machen soll (Abb. 10, S. 55). Die Leseansicht soll sich den Bedürfnissen des Nutzers stärker anpassen lassen. Beispielsweise sollen typographische Besonderheiten wie etwa die Umsetzung von Umlauten zugunsten besserer Lesbarkeit von der Anzeige ausgenommen oder in der Anzeige verändert werden können (ebd). Die für August 2011 angekündigte Leseansicht ist jedoch noch nicht in den produktiven Betrieb des DTA übergegangen (vgl. <http://www.deutschestextarchiv.de/planung/>). Die verschiedenen Leseansichten, die das DTA bislang zur Verfügung stellt und deren beschriebene Erweiterung unterstützen den geisteswissenschaftlichen Primitiv des Lesens, der die grundlegende Methode dieses Wissenschaftsparadigmas darstellt und dem bei der Entwicklung Virtueller Forschungsumgebungen stets besondere Beachtung in Bezug auf Effektivität und Nutzungskomfort zukommen muss.

The screenshot displays the DTA (Deutsches Textarchiv) web interface. On the left, a metadata sidebar provides details about the document: 'Vorlesungen über Esskunst' by 'Anthus, Antonius', published in 1838. The main content area shows the title page of the book, including the title 'Vorlesungen über Esskunst' by Antonius Anthus, the publisher 'Verlag von Otto Wigand' in Leipzig, and the year '1838'. Below the title page is the 'Vorwort' (Foreword) section. On the right, a 'Toolbox' offers various options for text presentation, such as 'Zeichendarstellung' (original vs. normalized), 'Hervorhebungen' (highlighting options like 'Überschriften', 'Zitate/Epigraphie', etc.), and an 'Inhaltsverzeichnis' (Table of Contents) with a depth of 2.

Abbildung 10: Zukünftige parametrisierbare Leseansicht

### 4.3.3 Webbasierte Korrekturumgebung

Weiterhin ist eine webbasierte Korrekturumgebung geplant, die den Vergleich von Imagescans und zugehörigen Volltexten vereinfacht, indem Bilder und Texte parallel angezeigt werden (Abb. 11, S. 56). Nutzer, die auf Fehler stoßen, können diese an die Datenbank melden und so direkt zur erweiterten Qualitätskontrolle beitragen. Durch diese Maßnahme kann die Qualität der Forschungsdaten für die Wissenschaftscommunity erhöht werden. Für das Melden von Fehlern ist eine differenzierte Fehlertypologie vorgesehen, mit der sich Fehler konsistent klassifizieren lassen, was ihre Korrektur durch die Datenbankmitarbeiter vereinfacht und beschleunigt. Fehler können damit bis auf die Ebene des einzelnen Zeichens identifiziert und den Kategorien Transkription, Darstellung/Anzeige, Auszeichnung

und Druck zugeordnet werden. Die Korrekturumgebung befindet sich momentan im institutsinternen Test.

The screenshot shows the DTA Fehlerdatenbank interface. At the top, it indicates the user is logged in as 'FrankWiegand' and shows the current page as 'Zufallsseite'. The main title is 'kurz\_sonnenwirth\_1855 (DON)'. The status bar shows 'Stand: Mon Apr 4 10:45:50 2011', 'Text: größer/kleiner', and '25 Fehler offen'. The interface is divided into several sections:
 

- Navigation:** Buttons for 'vorherige Seite', 'nächste Seite', and search options.
- Main Text Area:** Two columns of text from a historical document. The left column is a scan of a page with a page number '96'. The right column is a transcription of the same text with red annotations and a cursor.
- Right Sidebar:**
  - Fehler melden:** A list of error types such as 'Transkriptionsfehler', 'Auszeichnungsfehler (XML)', 'Darstellungsfehler (HTML)', etc.
  - Kommentar:** A text input field for adding notes.
  - Fehlerliste:** A table listing detected errors, including error IDs like '#3126 Auszeichnungsfehler (XML)' and '#3125 Auszeichnungsfehler (XML)', their dates, and the user 'FrankWiegand'.

Abbildung 11: Webbasierte Korrekturumgebung

Folgende weitere Funktionen soll die Korrekturumgebung dem Nutzer bieten (ebd.): Die Aufteilung der Text-Digitalisate in Seiten nach physischer Vorlage mit neuen Ansichten (Bildvorlage, TEI-XML, Text nach Transformation, automatisch erzeugte normierte Schreibung (CAB)); eine zeichengenaue (bildkoordinatengetreue) Erfassung der Fehlerstellen mit folgenden Hilfsfunktionen: eine Lupe für Bilddigitalisate, X-Path-Generierung, Anzeige potentieller Fehlerkandidaten, Statusmeldung 'wurde Korrektur gelesen' für jede korrigierte Seite; Benutzer- und Rechteverwaltung, Statistiken. In Planung befindet sich für das Deutsche Textarchiv eine Online-Korrekturfunktion, die direkt am Text bzw. seiner TEI-XML-Repräsentation ansetzt und eine Transkriptionsumgebung für Texte, die nicht das lateinische Alphabet verwenden, z.B. griechische, russische, hebräische Texte bietet. Für die enge Interaktion zwischen Nutzern und Deutschem Textarchiv bzw. BBAW-Mitarbeitern soll zukünftig eine „sichtbare Versionsgeschichte aller



Korrekturversionen“ (ebd.) und ein Ticketsystem für gemeldete Fehler implementiert werden.<sup>73</sup> Außerdem sollen die Rechte und Rollen für registrierte Nutzer feiner granuliert werden, um Diskretion und Vertrauen bei der Arbeit mit dem Deutschen Textarchiv zu fördern. Für einzelne Werke sollen Korrekturpatenschaften eingerichtet werden, bei denen einzelne Nutzer aus der Community die Korrektur eines bestimmten Werks verbindlich übernehmen. Die zunächst nur projektintern genutzte Korrekturumgebung DTA-Q (DTA-Q, 2012) ist mittlerweile auch für externe Nutzer freigeschaltet (derzeit 76 Nutzer). Zukünftig soll diese Umgebung zur Qualitätssicherung noch erweitert werden, um autorisierten Nutzern kleinere Eingriffe in die XML-Dateien der Texte zu erlauben und so z.B. Korrekturen von Druckfehlern, das Tagging von Eigennamen vorzunehmen oder editorische Kommentare beizufügen (mündl. Mitteilung vom 04.05.2012).

#### **4.3.4 Erweiterungsansätze**

Über diese institutionell angekündigten Weiterentwicklungen hinaus existiert noch Raum für Erweiterungen und Verbesserungen, die für die Arbeit mit dieser Virtuellen Forschungsumgebung wichtig und wünschenswert wären. Über die dem Deutschen Textarchiv zugrundeliegende Infrastruktur lässt sich keine Aussage treffen, da zu diesem Bereich des Archivs keine offizielle Dokumentation veröffentlicht wurde. Daher müssen sich die folgenden Anregungen auf den Nutzerbereich beschränken. Besonders der virtuelle Schreibtisch, der als „Mein Bereich“ bezeichnet ist, birgt noch Entwicklungspotenzial. Zwar ist die Anforderung nach einem virtuellen Schreibtisch erfüllt, aber der Schreibtisch sollte in seinen Funktionen erweitert werden, um komfortables Arbeiten zu ermöglichen. Beispielsweise ist es augenblicklich noch nicht möglich, sich unter „Mein Bereich“ alle angelegten Annotationen und Kategorien im Überblick anzeigen zu lassen oder die ausgewählten Texte als verlinkte Liste zu speichern. Daher muss der Anwender, sich die bearbeiteten Texte, Kategorien und gegebenenfalls auch Annotationen anderweitig speichern. Diese deutlich klaffende Lücke in der Unterstützung des Forschungsprozesses soll jedoch

---

<sup>73</sup> Über das geplante Ticketsystem sind noch keine Einzelheiten bekannt. Insgesamt können solche Systeme für Virtuelle Forschungsumgebungen nützlich sein, um das Voranschreiten des Projekts zu kontrollieren.

in absehbarer Zeit bei der geplanten Überarbeitung des gesamten Website geschlossen werden (mündl. Mitteilung vom 04.05.2012).

Darüber hinaus wäre die Möglichkeit, das selbst zusammengestellte, annotierte Korpus ganz oder partiell für alle oder bestimmte andere Nutzer der DTA zu veröffentlichen, vorteilhaft für die Lehre oder das Beforschen bestimmter Werkszusammenstellungen in der Gruppe. Dies würde den Primitiv der Kooperation/Kollaboration deutlich unterstützen. Wünschenswert wäre weiterhin die Gelegenheit, eine Literaturliste mit Sekundärliteratur zum eigenen Korpus bzw. zur eigenen Arbeit auf dem virtuellen Schreibtisch sowie zusätzliche Notizen oder eigene Artikel zu hinterlegen sowie die Verknüpfung des DTA mit verschiedenen Sekundärquellen wie Fachdatenbanken, Nachbarprojekten und elektronischer Fachliteratur und einem Open-Access-Server, der die Veröffentlichung eigener Forschungsergebnisse erlaubt. Auf diese Weise würde der wissenschaftliche Primitiv der Ablieferung/Veröffentlichung unterstützt, der im DTA noch nicht berücksichtigt ist.

Konzept und Umsetzung des Deutschen Textarchivs erweisen sich für die Linguistik, Literaturwissenschaft, Komparatistik und Geschichtsforschung als vielversprechend, auch wenn verschiedene Funktionen dieses im Aufbau befindlichen Projekts noch nicht überall im wünschenswerten Maße umgesetzt sind. Sicherlich kann das Deutsche Textarchiv nicht als Virtuelle Forschungsumgebung im eingangs definierten Sinn gelten. Allerdings ist es auch nicht erforderlich, für jede Wissenschaftsdisziplin eine eigene, von anderen abgegrenzte Virtuelle Forschungsumgebung zu konstruieren. Vielversprechender ist an dieser Stelle der Ansatz, verschiedenste Informationsressourcen und Werkzeuge im Rahmen einer generischen Forschungsinfrastruktur zu verknüpfen. Diese Infrastruktur wird erst in Verbindung mit solchen modularen Ressourcen, zu denen auch das Deutsche Textarchiv zählt, zu einer Virtuellen Forschungsumgebung im eigentlichen Sinne. Projekte wie das Deutsche Textarchiv zeigen dabei deutlich, was geisteswissenschaftlich Arbeitende inhaltlich und methodisch für ihre Forschungsaktivitäten benötigen und wo deutlicher Unterstützungsbedarf für den Übergang der Geisteswissenschaften in den digitalen Forschungsraum besteht. Gerade wo sich solche Lücken auftun, bieten sich für Bibliotheken neue

Gelegenheiten, diesen Bedarf durch ihre Expertise und Serviceorientierung zu decken. Mit diesen Möglichkeiten befasst sich das letzte Kapitel dieser Arbeit.

## **5. Virtuelle Forschungsumgebungen als Entwicklungsfeld für Bibliotheken**

### **5.1 Sammlungen, Daten, Wissen und Werkzeuge**

Die Vorzüge und Lücken des Deutschen Textarchivs zeigen nicht nur die Anforderungen wissenschaftlichen Arbeitens im digitalen Raum, sondern implizieren gleichzeitig Veränderungen, die Bibliotheken durchmachen müssen, um den *digital turn* in den Geisteswissenschaften angemessen zu begleiten.<sup>74</sup> Zu den angestammten Aufgaben von Bibliotheken als geisteswissenschaftliche Laboratorien gehört das Entwickeln, Erschließen und Zurverfügungstellen beforschbarer Primärinhalte in Absprache mit der Wissenschaft. Im Zeitalter der Digital Humanities müssen die Sammlungen von Forschungsbeständen ebenfalls digital angeboten werden. Somit ergibt sich für Bibliotheken und Forschung die gemeinsame Aufgabe, umfangreiche nachnutzbare digitale Sammlungen zu entwickeln (vgl. American Council of Learned Societies, 2006). Bibliotheken haben die Expertise, kooperativ qualitativ hochwertige Korpora zu erstellen, indem sie in Absprache mit den jeweiligen Fakultäten und Instituten relevante Bestände identifizieren und diese zu strukturierten Faksimiles digitalisieren und die Texte durch interoperable Formate wie etwa XML TEI P5 erschließen. Durch die Vielzahl gängiger quelloffen und frei zur Verfügung gestellter tagging-Werkzeuge, ist bibliotheksintern auch das part-of-speech-tagging und Lemmatisierung möglich, so dass die Volltexte auf Wortebene erschlossen und zur weiteren wissenschaftlichen Bearbeitung und Analyse (z.B. mit linguistischen Suchmaschinen) zur Verfügung gestellt werden können. Erst diese linguistische Tiefenerschließung macht aus Texten geisteswissenschaftlich verwertbare Primärforschungsdaten, die dem wissenschaftlichen Arbeitenden nutzen (vgl. Dudek, 2011, S. 44). Das bedeutet besonders für die Geisteswissenschaften die Digitalisierung historischer und zeitgenössischer Bestände zu qualitativ hochwertigen, semantisch analysierbaren Volltextkorpora, die in der historischen, diachronen und synchronen Forschung Verwendung finden. Diese Korpora müssen

---

<sup>74</sup> Vgl. auch Gradmann et al. (2008) zum Übergang vom Gutenberg- ins Turing-Universum digitalen Publizierens.

nicht nur mit semantischen Technologien (POS-tagger, KWIC-search, regular expressions, Strukturdaten etc.) durchsuchbar sein, sondern in ihrer Gesamtheit semantisch beschrieben werden, um im Semantic Web des WWW auffindbar zu sein.<sup>75</sup> Dazu gehört auch das Qualitätsmanagement für die Korpora. Die entstehenden Korpora können dann intern gehostet und mit anderen, auch dezentralen, Korpora verknüpft oder in diese externen Korpora eingespeist werden.<sup>76</sup>

In engem Zusammenhang mit der Erzeugung solcher Korpora steht deren Erschließung, durch die sie ersten in den Informationsraum der betreffenden Disziplin sicht- und nutzbar integriert werden. Diese Integration kann durch semantische Technologien erreicht werden. Durch die semantische Erfassung und Erschließung von Korpora und Werkzeugen sind diese im Internet besser auffindbar, wodurch die redundante, kostenintensive Neuentwicklung bestehender Objekte verhindert werden kann (vgl. American Council of Learned Societies, 2006). Dieses Erschließen von Sammlungen im digitalen semantischen Raum impliziert für Bibliotheken, dass sich für sie die Notwendigkeit ergibt, sich nicht mehr nur mit Containern bzw. Containerformaten für Wissensobjekte zu befassen, sondern die Inhalte dieser Container derart für wissenschaftliches Arbeiten nutzbar zu machen, dass gerade für Geistes- und Sozialwissenschaftler das notwendige stöbernde Suchen unterstützt und vereinfacht wird. Dieses extensive Konzept von Erschließungsaufgaben greift Fournier (2007) auf, indem er von Bibliotheken fordert, eine engere Bindung an Wissenschaftsdisziplinen zu entwickeln, um bessere fachspezifische Informationsdienstleistungen zu bieten. Eine solche Verbesserung könnten erreicht werden, indem Bibliotheken einerseits Wissen und Quellen im Netz finden, qualitativ bewerten und zugänglich machen und andererseits Dienste für virtuelle Zusammenarbeit in der Wissenschaft bieten. Bibliotheken müssten also den "Informationsraum qualifiziert [...] strukturieren" (Fournier, 2007, S. 165), indem sie

---

75 Welche hohen Anforderungen digitalisierte Texte erfüllen müssen, um für die Forschung tatsächlich nutzbar zu sein, verdeutlicht (Crane, 2006). Eine optische Zeichenerkennung (OCR) bei der Digitalisierung allein reicht dazu nicht aus. Erst eine Tiefenerschließung auf Wortebene durch Annotation von Wortarten, Eigennamen, Ortsnamen u.a. ermöglicht die Forschungsarbeit mit den Texten.

76 Das American Council of Learned Societies (2006) fordert hinsichtlich kommerziell erzeugter Sammlungen, die aus Beständen von Non-Profit-Organisationen wie Bibliotheken und Museen digitalisiert wurden, dass jeweils mindestens eine Version der Sammlung in einem nicht-proprietären Format an die bestandsbesitzenden Institutionen übergeben werden soll, die dort archiviert und lokal genutzt werden kann.

Standards für Dokumentstrukturen und Zugriffsprotokolle entwickeln, die Langzeitverfügbarkeit von elektronischen Dokumenten schaffen und Online-Informationen langfristig referenzierbar machen.<sup>77</sup> Dazu gehört auch die Entwicklung und Anwendung disziplinspezifischer Indizierungs- und Strukturierungsverfahren und die Indexierung elektronischer Informationsquellen für Suchmaschinen.<sup>78</sup> In der Konsequenz bedeutet der Imperativ der Strukturierung des Informationsraums auch die (Weiter-)Entwicklung und Verbesserung von Ontologien für eine Integration der Informationsobjekte ins Semantic Web.<sup>79</sup>

Die in den Sammlungen entstandenen Primärforschungsdaten bedürfen der differenzierten Verwaltung im Sinne eines Datenmanagements: Daten müssen sicher aufbewahrt und langzeitarchiviert werden, zugleich muss der dauerhafte Zugang mit differenzierten Berechtigungen gewährleistet sein. Die Arbeit mit solchen Primärdaten erfordert unter dem Gesichtspunkt wissenschaftlicher Qualitätskontrolle und Nachnutzung die differenzierte Zitierbarkeit der Daten. Damit stellt sich für Bibliotheken die Aufgabe, als Schnittstelle zwischen Forschung und technischer Infrastruktur zu fungieren (vgl. Neuroth et al., 2007, S. 273), die sich unter anderem in der Übernahme von differenzierten Aufgaben des Datenmanagements manifestiert.<sup>80</sup> In diesem Zusammenhang können und sollten sich Bibliotheken auch an der Entwicklung und Pflege Offener Standards und Analysewerkzeuge beteiligen, um die Langzeitverfügbarkeit und problemlosen Datenaustausch zu gewährleisten (vgl. American Council of Learned Societies, 2006).

---

77 Auch die Entwicklung von Möglichkeiten zur Tiefenzitation einzelner Datensätze aus Korpora oder Internetquellen wie Webseiten sollte Teil dieser Aufgaben sein.

78 Die Indexierung qualitativ hochwertiger Wissensquellen aller Art im Netz erübrigt m. E. die von Fournier (2007; S. 165) geforderten zertifizierten Linksammlungen. Datenbanken, Projekte und andere Quellen (z.B. eine extensive Sammlung relevanter Sprachkorpora wie die der Humboldt Universität zu Berlin (2011)) werden durch die Indexierung für Literatursuchmaschinen wie ELIB, vufind, Touchpoint, Primo, Summon u.a. besser auffindbar, etwa durch die Anreicherung der Metadaten durch Schlagwörter oder - noch eleganter - durch die Einordnung in eine Ontologie, die von Semantic Web Tools ausgewertet werden kann.

79 Palmer et al. (2009, S. 37–41) zeigen, welche weiteren konkreten Bibliotheksdienstleistungen die von ihnen rekonstruierten wissenschaftlichen Primitive unterstützen können.

80 Einen ersten Vorstoß in diesem Bereich macht die Staats- und Universitätsbibliothek Bremen mit dem Projekt Qualiservice (Qualiservice, 2012), bei dem qualitative Daten aus der Sozialwissenschaft durch die Bibliothek strukturiert und erschlossen werden.

## **5.2 Generische Virtuelle Forschungsumgebungen und Forschungspartnerschaft**

Die Darstellung von Korpora und anderen Forschungsdaten kann in einem Internetportal der Bibliothek erfolgen, wie etwa im Fall der „Digitalen Sammlungen“ der Staats- und Universitätsbibliothek Bremen, die ein digitales Korpus von Bremensien und zukünftig auch ein Zeitungskorpus zur Verfügung stellt. Von dieser Darstellung der Sammlungen und Verwaltung von Forschungsdaten zum Anbieten von Virtuellen Forschungsumgebungen, in denen Forschung – auch kollaborativ – direkt betrieben werden kann, ist der Schritt nicht weit. Auch Fournier (2007) sieht die Aufgabe von Bibliotheken in der Unterstützung und Weiterentwicklung generischer Virtueller Forschungsumgebungen in Zusammenarbeit mit hochschulbezogenen Infrastruktureinrichtungen wie Rechen- und Medienzentren. Besonders das Projekt TextGrid illustriert in diesem Zusammenhang, dass Bibliotheken großes Entfaltungspotential in der Entwicklung Virtueller Forschungsumgebungen finden. Etwa liegt die Leitung des TextGrid-Projekts bei der Staats- und Universitätsbibliothek Göttingen, die unter anderem aufgrund dieser und ähnlicher Aktivitäten als erste deutsche Bibliothek eine Abteilung für Forschung und Entwicklung eingerichtet hat. So wird nicht nur die Volltext-Digitalisierung analoger Ressourcen für Unternehmungen dieser Art von Bibliotheken übernommen, sondern auch das Hosting und Ressourcenmanagement im Bereich der Forschungsinfrastruktur und die Entwicklung von Forschungswerkzeugen. Bibliotheken können folglich den Grundstein zu hochschulübergreifenden Forschungsportalen legen, die zunehmend mit Primärforschungsdaten angereichert und wissenschaftlich genutzt werden. Im Hinblick auf solche Forschungsportale, die in Funktion und Aufbau generische Virtuelle Forschungsumgebungen sein müssen, ist es wichtig, Repositorien mit verschiedenen Datenkorpora zu verknüpfen, um das Suchen, Finden und Vergleichen geeigneter Forschungsdaten besonders für die textorientierten Wissenschaften zu unterstützen. Eine sinnvolle Ergänzung dieser Umgebungen sind auch Verknüpfungen mit Fachdatenbanken (z.B. Kindlers Literaturlexikon, BLL etc.) und E-Bookbeständen (z.B. für Literaturgeschichten, Werkinterpretationen etc.), die die Beurteilung von Texten bei der Auswahl des Forschungskorpus vereinfachen und damit den Aufbau individueller

Forschungskorpora unterstützen. Abbildung 12 (S. 64) stellt schematisch dar, welche Elemente eine für alle – nicht nur geistes- und sozialwissenschaftliche – Disziplinen generische und dennoch spezifische Virtuelle Forschungsumgebung beinhalten sollte. Das DTA bietet keine Möglichkeit zur Publikation des selbstzusammengestellten Korpus oder darauf basierender Veröffentlichungen. Da in den Geisteswissenschaften Wissen jedoch nicht über die von Knorr-Cetina (2002) für die Naturwissenschaften beschriebenen sozialen Netzwerke oder zeitnah erscheinende Zeitschriftenpublikationen verbreitet wird, ist es zur Beschleunigung des Publikationsprozesses wichtig, dass Virtuelle Forschungsumgebungen diesen aktiv unterstützen. Bibliotheken können diese Möglichkeit bieten. Eine solche Unterstützung lässt sich durch Open Access-Repositoryn und -Services umsetzen. Um wissenschaftlich Etablierten wie Nachwuchswissenschaftlern die Veröffentlichung ihrer Informationsobjekte und Resultate auf schnelle und komfortable Weise zu ermöglichen und sie gleichzeitig für die Fachöffentlichkeit gut zugänglich zu machen, können Bibliotheken sowohl die Open Access-Infrastruktur als auch das zugehörige Veröffentlichungsmanagement stellen. Fournier (2007) schlägt daher vor, bei der Weiterentwicklung von Bibliotheken gezielt auf das Publikationsbedürfnis von Wissenschaftlern einzugehen und sie bei der Bereitstellung und Verbreitung ihrer neuerzeugten Inhalte zu unterstützen. Diese Unterstützung verteilt sich dieser Argumentation nach auf vier Servicebereiche. Den ersten bilden differenzierte Publikationsservices, die die Klärung von Veröffentlichungsrechten, die Qualitätssicherung von Layout und Usability der Publikationen, die professionelle Erstellung von Metadaten durch inhaltliche und klassifikatorische Erschließung und den Upload auf den Publikationsserver beinhalten und somit verlegerische Aufgaben an die Bibliothek herantragen. Den zweiten Bereich bilden technische Dienstleistungen wie die Garantie der Langzeitverfügbarkeit und Tiefenzitation (z.B. aus großen Datenkorpora oder von Webseiten) sowie die Verknüpfung von Zitaten und ihrer Quellen in der elektronischen Veröffentlichung selbst. Der dritte Bereich umfasst die Reputationsgenese für Open Access-Publikationen, indem Bibliotheken neue Werkzeuge zur Messung der inhaltlichen Qualität von Veröffentlichungen entwickeln

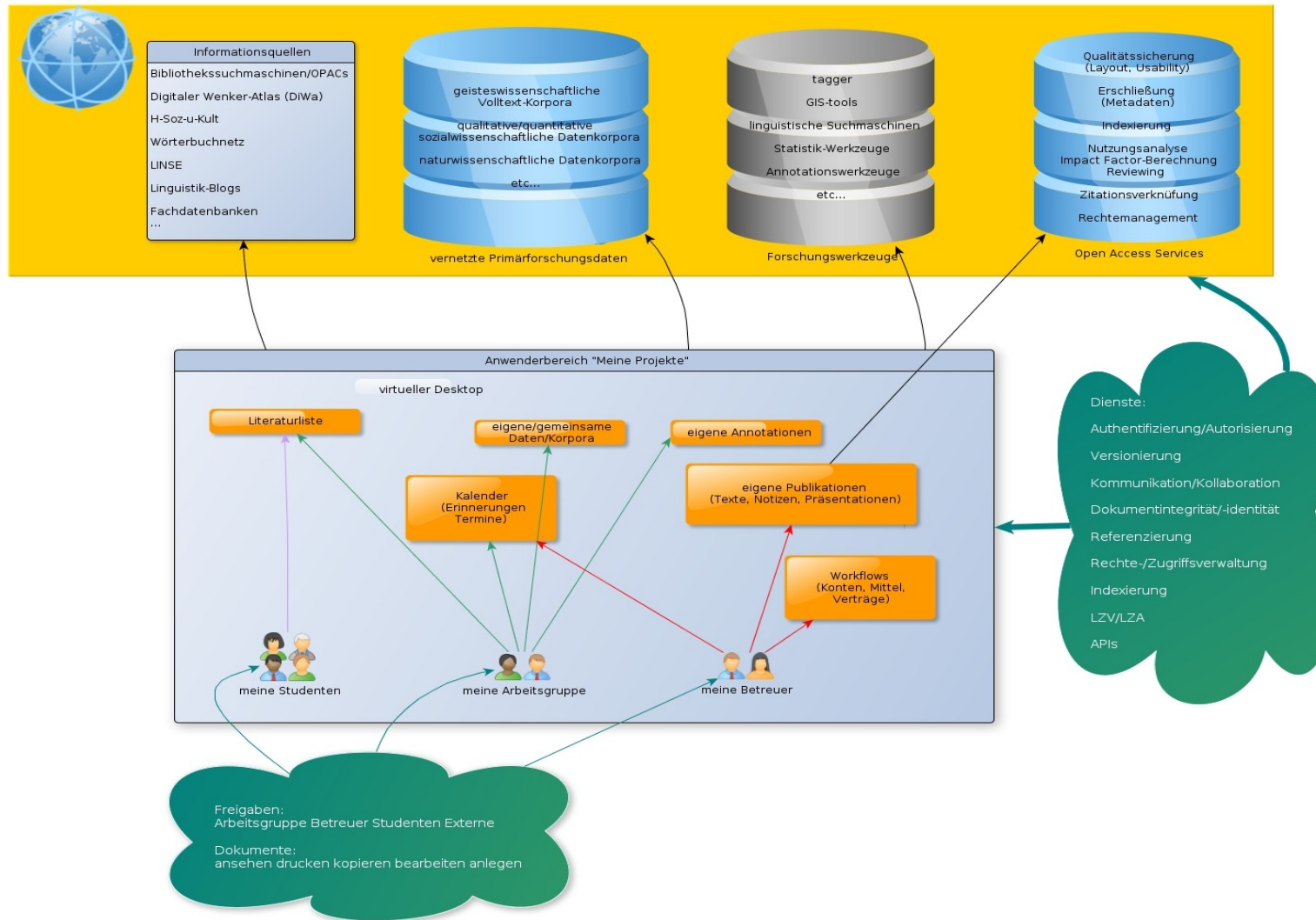


Abbildung 12: Generische Virtuelle Forschungsumgebung



und anwenden. Dazu gehören vor allem neue Tools zur Nutzungsanalyse von Publikationen und damit auch die Konzeptionierung alternativer, vertrauenswürdiger Verfahren zur Messung des Impact Factors von Publikationen und Autoren.<sup>81</sup> Der vierte Bereich ist geprägt durch die strukturierte Erweiterung des Informationsraums durch die Erschließung flüchtiger Kommunikationsformen. Wissenschaftsblogs wie etwa das E-Learning Weblog des Leibniz-Zentrums für Psychologische Information und Dokumentation (ZPID) (vgl. Leibniz-Zentrum für Psychologische Information und Dokumentation, 2012) sind verlässliche und wissenschaftlich relevante Quellen, die es zu erschließen gilt. Ein Beispiel für eine solche Integration, die eine Basis für die semantische Recherche im Internet bietet, ist die Verknüpfung von Personennennungen in Wikipedia-Artikeln mit der Personennamendatei (PND) der Deutschen Nationalbibliothek. Aus diesen vier Bereichen leitet Fournier (2007) konkrete Aufgaben für Bibliotheken ab, die sich jedoch über den Publikationsbereich hinaus auch auf die Entwicklung und Bereitstellung von Virtuellen Forschungsumgebungen erstrecken: Bibliotheken müssen Autorisierungs- und Authentifizierungsverfahren für die vertrauenswürdige Nutzung von Virtuellen Forschungsumgebungen bieten, Standards und Verfahren für die Langzeitverfügbarkeit elektronischer Publikationen bieten und dazu entsprechende Repositorien für Daten und Publikationen aufbauen und vernetzen. Weiterhin ist es wichtig, Werkzeuge wie Infrastruktur zum Finden, Strukturieren und Erschließen von Informationen sowie für fachliche Kooperation und Kollaboration bereitzustellen.

Virtuelle Forschungsumgebungen als neues Aufgabenfeld für Bibliotheken bedeuten eine Veränderung ihrer Rolle und Position im Wissenschaftsfeld weg vom nur zuarbeitenden Dienstleister zu einem aktiven Partner der Forschung, der neue Technologien entwickelt, bereitstellt und vermittelt, die Wissensflut im Netz durch ihre inhaltlich-semantische Aufbereitung beherrschbar macht und den Prozess der Verbreitung neuen Wissens vorantreibt.

---

<sup>81</sup> Diesen Schritt weg von Containern, hin zu den Inhalten von Wissen und Forschung vertreten auch Gradmann (2009) und Linten et al. (2011).

### 5.3 Die Zukunft des Fachreferats

Die beschriebenen Entwicklungen wirken sich auch auf das Berufsbild des Fachreferenten aus. Der Fachreferent der Zukunft sieht sich im Zeitalter des Übergangs von der analogen in die digital-analoge Welt einem erweiterten Aufgabenfeld gegenüber. Die Produktion von Text im Allgemeinen und von wissenschaftlichen Texten im Speziellen hat sich in Zeiten des Web 2.0, das von den dynamischen Beiträgen vieler Autoren, die sich nicht einmal unbedingt als solche sehen, deutlich verändert. Diese und damit zusammenhängende Veränderungen werden momentan unter dem Begriff *Gutenberg-Klammer* im Rahmen verschiedener Forschungsvorhaben in den USA und Dänemark, wo dieser Begriff geprägt wurde, untersucht.<sup>82</sup> Text ist unter der Bedingung dynamischer, informeller Autorschaft im Netz vom Produkt zum Prozess geworden (vgl. University of Southern Denmark, 2012). Text wird nicht mehr nur von dezidierten Autoren produziert und durch seinen Druck für die Ewigkeit fixiert, sondern entsteht heute auch durch die Arbeit eines oder mehrerer 'unsichtbarer', informeller Autoren, wird gelöscht, revidiert, editiert, ohne dass eine 'offizielle' (d.h. formelle) Fertigstellung wie sie der Druck erfordert, überhaupt angestrebt sein muss.<sup>83</sup> Aus dem neuen prozessualen Charakter der Textproduktion im Netz, dem also auch ein neues Konzept von Autorschaft zugrundeliegt, ergibt sich auch eine Konsequenz für den Fachreferenten der Zukunft: auch die prozessualen Texte flüchtiger Kommunikationsformen mit informeller, anonymer Autorschaft wie Blogs oder Wikis müssen auf ihre Zuverlässigkeit und Verwertbarkeit als Fachtexte geprüft und als solche erschlossen werden, um das darin gespeicherte Wissen zugänglich zu machen. Darüber hinaus leitet sich daraus die Möglichkeit ab, solche Formen informeller Wissensgenerierung für das eigene Fachgebiet aktiv mitzugestalten. So kann das bibliothekarische und fachwissenschaftliche Potential des Fachreferenten vollständig zur Erschließung von Information genutzt werden. Auch hier macht sich der Paradigmenwechsel von

---

82 Eine kurze Zusammenfassung des Konzepts bietet Markdot (2012) und ist gleichzeitig ein prägnantes Beispiel für die neue Art dynamischer Textproduktion.

83 Dieses Konzept anonymer oder anonymisierender Autorschaft ist typisch bis ins Mittelalter hinein. Von der Antike bis in die Zeit vor dem Buchdruck nahm sich der Autor oder Übersetzer eines Textes tendenziell persönlich zurück und schrieb die intellektuelle Leistung häufig dem eigenen Lehrer (z.B. Sokrates im Falle Platons) oder einem gefühlt berühmteren Autor zu (z.B. Wolfram gegenüber Chrétien de Troyes). Auch den Autoren eines Werkes gar nicht zu nennen, war durchaus üblich.

Containern zu Inhalten bemerkbar. Möglicherweise deutet sich damit eine gewisse Rückbesinnung auf die ursprüngliche Funktion von Bibliotheken als Ort der Generierung und Aufbewahrung von Wissen an.

Für den Fachreferenten der Zukunft bietet sich ein heterogenes, ausgeweitetes Berufsbild: er findet, bewertet und erschließt Informationsquellen im Internet (dazu gehören Daten, Sekundärliteraturquellen, Projekte, Blogs, Datenbanken, E-Books etc.) und integriert sie als Quellen in Virtuelle Forschungsumgebungen. Dabei beteiligt er sich an der Strukturierung des Wissensraums seines Fachgebiets. Er berät Wissenschaftler beim elektronischen Publizieren (auch in rechtlicher Hinsicht) und bezüglich geeigneter Werkzeuge für ihre Forschungsansätze und Bedürfnisse. (Dies erfordert den engen Kontakt zum Seminar, zur Fakultät und kooperierenden Instituten, die er betreut.) Neben der Vermittlung von Informationskompetenz im Umgang mit den klassischen fachwissenschaftlichen Datenbanken schult er den Umgang mit der hochschuleigenen Virtuellen Forschungsumgebung sowie grundlegende Methoden.

## **6. Bibliotheken und Virtuelle Forschungsumgebungen**

Das Deutsche Textarchiv zeigt, dass Informationsressourcen, die technisch und semantisch hochwertig aufbereitet wurden, zwar unverzichtbar für das wissenschaftliche Arbeiten von Philologen und anderen Geisteswissenschaftlern sind, aber nur einen geringen Kernbereich des Forschungsprozesses abbilden. Eine komfortable Benutzung, die das Speichern, Auswerten und Veröffentlichen der Daten und Ergebnisse sowie Zusammenarbeit erlaubt, ist noch nicht möglich. Hier besteht ein umfangreicher Weiterentwicklungsbedarf, dessen Ziel es sein sollte, die Unterstützung wissenschaftlicher Primitive deutlich zu verbessern. Es stellt sich die Frage, ob dieser Verbesserungsbedarf tatsächlich beim Deutschen Textarchiv, das in erster Linie nur Informationsressourcen und Werkzeuge bieten soll, zu sehen ist. Eine Alternative zum Ausbau ausgefeilter digitaler Korpora zu kostspieligen, singulären Virtuellen Forschungsumgebungen (die durch ihr Stand-alone-Konzept auch Gefahr laufen, als Insellösungen zu enden), bietet die Integration solcher Forschungsressourcen in ein generisches technisches Framework, das von

Bibliotheken in Zusammenarbeit mit universitären Infrastruktureinrichtungen hochschul- oder sogar deutschlandweit angeboten wird. Dabei gilt es, die zentralen Anforderungen an Virtuelle Forschungsumgebungen, Generizität, Modularität, Interoperabilität, Abbildung wissenschaftlicher Primitive, Integration von Informationsressourcen und differenziertes Rechtemanagement, im Rahmen einer angemessenen technischen Infrastruktur zu berücksichtigen (s. Tab. 2, S. 37).

Das Übernehmen dieser Aufgabe ist für Bibliotheken auch durch ein verändertes Selbstverständnis gekennzeichnet. Im Digitalzeitalter muss die Aufmerksamkeit wieder den Inhalten geistigen Schaffens gelten, statt ihren Formaten. Gradmann (2009, S. 4) formuliert dazu:

"Angesichts von Milliarden und Abermilliarden persistent verfügbarer digitaler Informationsobjekte wird es ja schlicht nicht mehr möglich sein, allein durch „Lesen“ und menschliche, intellektuelle Aktivität diese Informationsmassen gedanklich zu durchdringen – zumindest so weit, dass immerhin eine im jeweils gegebenen Kontext angemessene Auswahl dieser Objekte möglich würde!"

Zentrale Aufgabe für Bibliotheken muss also neben der (Retro-)Digitalisierung von Sammlungen zu qualitativ hochwertigen, leicht prozessierbaren Korpora (vgl. Meyer, 2011, S. 43) die ebenso hochwertige Erschließung des darin gespeicherten Wissens sein. Bibliotheken müssen sich auch maßgeblich an der Vernetzung und Implementierung solcher Korpora in Virtuellen Forschungsumgebungen beteiligen, um Forschung im digitalen Raum zu erleichtern. Damit kommt auf Bibliotheken auch die Aufgabe zu, neue Werkzeuge zur Erschließung, Analyse und Weiterbearbeitung dieser Informationsobjekte zu entwickeln (vgl. Gradmann, 2009) und die dabei entstehenden Primärforschungsdaten adäquat zu verwalten, zu archivieren und zu referenzieren.

Eine weitere Herausforderung besteht für Bibliotheken in der umfassenden Unterstützung des wissenschaftlichen Publikationsprozesses in all seinen Facetten vom Layout bis hin zur bibliometrischen Qualitätssicherung (vgl. Short, 2005, S. 24). Was bereits Mittler (1994) mit den „drei Revolutionen des EDV-Zeitalters“ für Bibliotheken andeutet, bedeutet einen Funktions- bzw. Anforderungszugewinn für Bibliotheken. Den Medienmix der hybriden Bibliothek zu beherrschen und sich zum kompetenten Führer darin zu entwickeln, bedeutet in Bezug auf das digitale

Forschungszeitalter der Virtuellen Forschungsumgebungen, zum kompetenten Berater der Wissenschaft für technische Methoden und Werkzeuge der Digital Humanities zu werden. Über diese neuen Chancen entwickeln sich Bibliotheken im digitalen Zeitalter weg vom zuarbeitenden Informationsdienstleistung hin zum Forschungspartner, der Projekte in Zusammenarbeit mit der Wissenschaft plant und umsetzt. Das ist nur möglich, wenn sich Bibliotheken und Bibliothekare über ihre Expertise bewusst sind und sie auch nach außen signalisieren. Die Notwendigkeit technischer und fachwissenschaftlicher Expertise erfordert eine angemessene Orientierung der Ausbildung und Personalentwicklung: wissenschaftliche Bibliothekare müssen heute ein hohes Maß an Kompetenz in Bezug auf relevante Themen der Informationstechnologie mitbringen, das weit über die Anwendung von Office-Programmen hinausgeht und in der bibliothekarischen Ausbildung entwickelt, gefördert und gefordert wird. Die Möglichkeiten der Forschung im Cyberspace fordern Bibliotheken heute dazu auf, 'mittendrin' statt nur dabei zu sein und die Entwicklung dieses digitalen Raums aktiv mitzugestalten.

## Literaturverzeichnis

- American Council of Learned Societies (2006). *Our Cultural Commonwealth: The report of the American Council of Learned Societies Commission on cyberinfrastructure for the humanities and social sciences*.  
URL:[http://www.acls.org/uploadedFiles/Publications/Programs/Our\\_Cultural\\_Commonwealth.pdf](http://www.acls.org/uploadedFiles/Publications/Programs/Our_Cultural_Commonwealth.pdf) [Zuletzt geprüft am 05.03.2012].
- Aschenbrenner, A.; Blanke, T.; Dunn, S.; Kerzel, M.; Rapp, A. und Zielinski, A. (2007). *Von e-Science zu e-Humanities – Digital vernetzte Wissenschaft als neuer Arbeits- und Kreativbereich für Kunst und Kultur*. BIBLIOTHEK Forschung und Praxis (31), Nr. 1, (S. 11–21) [Zuletzt geprüft am 22.08.2011].
- Bakewell, E.; Beeman, W. und Mc Michael Reese, C. (1988). *Object, image, inquiry. The art historian at work*, Santa Monica California: Getty Art History Information Program.
- Bernardou, A.; Constantopoulos, P.; Dallas, C. und Gavrilis, D. (2010). *Understanding the Information Requirements of Arts and Humanities Scholarship*. The International Journal of Digital Curation, Nr. 5,1, (S. 18–33). URL:  
<http://ijdc.net/index.php/ijdc/> [Zuletzt geprüft am 05.03.2012].
- Berners-Lee, T. (2009). *Linked Data - Design Issues*.  
URL:<http://www.w3.org/DesignIssues/LinkedData.html> [Zuletzt geprüft am 06.03.2012].
- Berners-Lee, T.; Hendler, J. und Lassila, O. (2001). *The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities*. Scientific American (284), Nr. 5, (S. 29–37).  
URL:<http://www.med.nyu.edu/research/pdf/mainim01-1484312.pdf>.
- Blanke, T. und Hedges, M. (2011). *Scholarly primitives: Building institutional infrastructure for humanities e-Science*. ScienceDirect - Future Generation Computer Systems, (S. 1–8). URL:<http://dx.doi.org/10.1016/j.future.2011.06.006> [Zuletzt geprüft am 24.10.2011].
- Blumauer, A. und Pellegrini, T. (2006). *Semantic Web und semantische Technologien: Zentrale Begriffe und Unterscheidungen*. In Pellegrini, T. und Blumauer, A. (Hrsg.), *Semantic Web. Wege zur vernetzten Wissensgesellschaft*, (S. 9–25), Berlin, Heidelberg: Springer Berlin Heidelberg.
- Bohnsack, R. (2007). *Die dokumentarische Methode und ihre Forschungspraxis: Grundlagen qualitativer Sozialforschung*, 2. Auflage, Wiesbaden: VS, Verl. für Sozialwiss.
- Bourdieu, P. (2002). *Homo academicus*, Suhrkamp Taschenbuch Wissenschaft (Band 1002), 1. Auflage, Frankfurt am Main: Suhrkamp.

- Bourdieu, P. (2008). *Sozialer Sinn: Kritik der theoretischen Vernunft*, Suhrkamp-Taschenbuch Wissenschaft (Band 1066), 1. Auflage, Frankfurt am Main: Suhrkamp.
- Bourdieu, P. und Schwibs, B. (2005). *Die feinen Unterschiede: Kritik der gesellschaftlichen Urteilskraft*, Suhrkamp Taschenbuch Wissenschaft (Band 658), 16. Auflage, Frankfurt am Main: Suhrkamp.
- Brockman, W. S.; Neumann, L.; Palmer, C. L. und Tidline, T. J. (2001). *Scholarly Work in the Humanities and the Evolving Information Environment*. URL: <http://old.diglib.org/pubs/dlf095/dlf095.pdf> [Zuletzt geprüft am 05.03.2012].
- Butler, D. (2003). *The Grid: Tomorrow's computing today*. Nature (422), Nr. 6934, (S. 799–800).
- Carr, L.; Miles-Board, T.; Wills, G.; Power, G.; Bailey, C.; Hall, W. und Grange, S. *Extending the role of the digital library*. In Association for Computing Machinery (Hrsg.). (2004), *Proceedings of the Fifteenth ACM Conference on Hypertext & Hypermedia - HYPERTEXT '04*, New York, (S. 12–21).
- CLARIN (2012). *CLARIN - Common Language Resources and Technology Infrastructure*. URL: <http://www.clarin.eu/> [Zuletzt geprüft am 22.03.2012].
- Clarke, C. und Greig, F. (2009). *Case Study: A Linked Open Data Resource List Management Tool for Undergraduate Students*. URL: <http://www.w3.org/2001/sw/sweo/public/UseCases/Talis/> [Zuletzt geprüft am 06.03.2012].
- Clio-online (2012). *Geschichte und Internet - Clio-online*. URL: <http://www.clio-online.de/> [Zuletzt geprüft am 09.05.2012].
- Crane, D. (1972). *Invisible colleges: diffusion of knowledge in scientific communities*, Chicago [u.a.]: University of Chicago Press.
- Crane, G. (2006). *What Do You Do with a Million Books?* D-Lib Magazine (12), Nr. 3. URL: <http://www.dlib.org/dlib/march06/crane/03crane.html> [Zuletzt geprüft am 16.04.2012].
- DARIAH (2012). *DARIAH - Digital Research Infrastructure for the Arts and Humanities*. URL: <http://www.dariah.eu/> [Zuletzt geprüft am 22.03.2012].
- Deutsche Forschungsgemeinschaft (2006). *Wissenschaftliche Literaturversorgungs- und Informationssysteme: Schwerpunkte der Förderung bis 2015. DFG Positionspapier*. URL: <http://www.dfg.de/download/pdf/foerderung/programme/lis/positionspapier.pdf> [Zuletzt geprüft am 15.11.2011].
- Deutsche Forschungsgemeinschaft (2009). *Merkblatt. Service-Projekte zu Informationsmanagement und Informationsinfrastruktur in Sonderforschungsbereichen INF*.

URL:[http://www.dfg.de/download/programme/sonderforschungsbereiche/allgemeine\\_hinweise/60\\_06/60\\_06.pdf](http://www.dfg.de/download/programme/sonderforschungsbereiche/allgemeine_hinweise/60_06/60_06.pdf) [Zuletzt geprüft am 15.11.2011].

Deutsche Forschungsgemeinschaft (2010). *Aufforderung zur Antragstellung. Informationsmanagement. Ausschreibung "Informationsinfrastrukturen für Forschungsdaten"* (28.04.2010).

URL:[http://www.dfg.de/download/pdf/foerderung/programme/lis/ausschreibung\\_forschungsdaten\\_1001.pdf](http://www.dfg.de/download/pdf/foerderung/programme/lis/ausschreibung_forschungsdaten_1001.pdf) [Zuletzt geprüft am 16.11.2011].

Deutsche Forschungsgemeinschaft (2011a). *DFG - Deutsche Forschungsgemeinschaft - Förderbereich: Informationsmanagement.*

URL:[http://www.dfg.de/foerderung/programme/infrastruktur/lis/lis\\_themenschwerpunkt\\_digitale\\_information/lis\\_informationsmanagement/index.html](http://www.dfg.de/foerderung/programme/infrastruktur/lis/lis_themenschwerpunkt_digitale_information/lis_informationsmanagement/index.html) [Zuletzt geprüft am 07.02.2012].

Deutsche Forschungsgemeinschaft (2011b). *Merkblatt Virtuelle Forschungsumgebungen.*

URL:[http://www.dfg.de/download/formulare/12\\_12/12\\_12.pdf](http://www.dfg.de/download/formulare/12_12/12_12.pdf) [Zuletzt geprüft am 29.09.2011].

Deutsche Forschungsgemeinschaft (2011c). *Merkblatt Werkzeuge und Verfahren des wissenschaftlichen Informationsmanagements.*

URL:[http://www.dfg.de/formulare/12\\_13/12\\_13.pdf](http://www.dfg.de/formulare/12_13/12_13.pdf) [Zuletzt geprüft am 16.11.2011].

Deutsche Forschungsgemeinschaft (2011d). *Programm „Virtuelle Forschungsumgebungen“.*

URL:[http://www.dfg.de/foerderung/programme/infrastruktur/lis/lis\\_themenschwerpunkt\\_digitale\\_information/lis\\_informationsmanagement/index.html#micro161625](http://www.dfg.de/foerderung/programme/infrastruktur/lis/lis_themenschwerpunkt_digitale_information/lis_informationsmanagement/index.html#micro161625) [Zuletzt geprüft am 29.09.2011].

DiWA (2012). *Digitaler Wenker-Atlas (DiWA)*. URL: <http://www.3.diwa.info/> [Zuletzt geprüft am 17.04.2012].

DTA (2011). *Deutsches Textarchiv (DTA)*. URL: <http://www.deutschestextarchiv.de/> [Zuletzt geprüft am 07.11.2011].

DTA-E (2012). *Deutsches Textarchiv Ergänzungskorpus (DTA-E)*.

URL:<http://www.deutschestextarchiv.de/dtae> [Zuletzt geprüft am 09.05.2012].

DTA-Q (2012). *Deutsches Textarchiv Qualitätssicherung (DTA-Q)*.

URL:<http://www.deutschestextarchiv.de/dtaq> [Zuletzt geprüft am 09.05.2012].

Dudek, S. (2011). *Schöne Literatur binär kodiert : Die Veränderungen des Text- und Dokumentbegriffs am Beispiel digitaler Belletristik und die neue Rolle von Bibliotheken.* URL:<http://edoc.hu-berlin.de/series/berliner-handreichungen/2011-290/PDF/290.pdf> [Zuletzt geprüft am 03.11.2011].



- DWDS (2012). *Digitale Wörterbuch der deutschen Sprache des 20. Jahrhunderts* (DWDS). URL: <http://www.dwds.de/> [Zuletzt geprüft am 20.04.2012].
- eAQUA (2012). eAQUA. URL: <http://www.eaqua.net/> [Zuletzt geprüft am 19.04.2012].
- Eckert, K. (2011). *SKOS: eine Sprache für die Übertragung von Thesauri ins Semantic Web*. URL: <http://metadaten-twr.org/2011/01/19/skos-simple-knowledge-organisation-system/> [Zuletzt geprüft am 01.02.2012].
- Eckert, K.; Hänger, C. und Niemann, C. (2009). *Tagging and automation: challenges and opportunities for academic libraries*. *Library Hi Tech* (27), Nr. 4, (S. 557–569) [Zuletzt geprüft am 01.02.2012].
- Eckert, K.; Pfeffer, M. und Stuckenschmidt, H. (2008). *Assessing thesaurus-based annotations for semantic search applications*. *International Journal of Metadata, Semantics and Ontologies* (3), Nr. 1, (S. 53).
- Egan, M. E. (Hrsg.) (1954). *The communication of specialized information.*, Chicago: American Library Association for the University of Chicago Graduate Library School.
- Engels, R. und Tønnesen, J. R. (08.07.2009). *Case Study: A Digital Music Archive (DMA) for the Norwegian National Broadcaster (NRK) using Semantic Web techniques*. URL: <http://www.w3.org/2001/sw/sweo/public/UseCases/NRK/> [Zuletzt geprüft am 06.03.2012].
- Foster, I. und Kesselman, C. (1999). *The Grid: Blueprint for a New Computing Infrastructure*, San Francisco, Calif: Kaufmann.
- Foster, I.; Kesselman, C. und Tuecke, S. (2001). *The Anatomy of the Grid. Enabling Scalable Virtual Organizations*. URL: <http://www.globus.org/alliance/publications/papers/anatomy.pdf> [Zuletzt geprüft am 11.01.2012].
- Foster, I. (2002). *What is the Grid? A Three Point Checklist*. URL: <http://www.mcs.anl.gov/~itf/Articles/WhatIsTheGrid.pdf> [Zuletzt geprüft am 11.01.2012].
- Fournier, J. (2007). *Elektronisches Publizieren im Wandel. Zur Akzeptanz neuer Formen wissenschaftlicher Kommunikation*. In Lülfi, D.; Benkert, H. und Siebert, S. (Hrsg.), *95. Deutscher Bibliothekartag in Dresden 2006*, (S. 163–172), Frankfurt am Main: Klostermann.
- Fraser, M. (2005). *Virtual Research Environments: Overview and Activity*. *Ariadne*, Nr. 44. URL: <http://www.ariadne.ac.uk/issue44/fraser/>.
- Gould, C. C. (1988). *Information needs in the humanities. An assessment*, Stanford, California: Research Libraries Group.

- Gould, C. C. und Handler, M. (1989). *Information needs in the social sciences. An assessment. Prepared for the Program for Research Information Management of The Research Libraries Group, Inc.*, Mountain view California: Research Libraries Group. URL: H61.3 G696 1989.
- Gould, C. C. und Pearce, K. (1991). *Information needs in the sciences. An assessment*, Mountain view California: Research Libraries Group. URL: <http://worldcatlibraries.org/wcpa/oclc/474621045>.
- Gradmann, S. (2009). *Signal. Information. Zeichen. Zu den Bedingungen des Verstehens in semantischen Netzen. Antrittsvorlesung 28. Oktober 2008*. Libreas, Nr. 14, (S. 44–52). URL: <http://edoc.hu-berlin.de/humboldt-vl/157/gradmann-stefan-3/PDF/gradmann.pdf> [Zuletzt geprüft am 15.11.2011].
- Gregory, I. N. und Hardie, A. (2011). *Visual GSting: bringing together corpus linguistics and Geographical Information Systems*. Literary and Linguistic Computing (26), Nr. 3, (S. 297–314). URL: <http://llc.oxfordjournals.org/content/26/3/297.full.pdf#page=1&view=FitH>.
- Grote, T. (2011). *Are Intellectual Monopolies Morally Justified?*, Institut für Philosophie: Universität Potsdam, Potsdam [Zuletzt geprüft am 19.03.2012].
- Heinrich-Heine-Portal (2011). *Heinrich-Heine-Portal*. URL: <http://www.hhp.uni-trier.de/> [Zuletzt geprüft am 22.08.2011].
- Henn-Memmesheimer, B. (2006). *Wortgeschichten: Driften im semantischen Raum*. In Eitelmann, M. und Stritzke, N. (Hrsg.), *Ex praeteritis praesentia*, (S. 43–68), Heidelberg: Winter. URL: [http://germanistik.uni-mannheim.de/abteilungen/germanistische\\_linguistik/prof\\_dr\\_beate\\_henn\\_memmesheimer/publikationen\\_henn\\_memmesheimer/wortgeschichten\\_ultima\\_mit\\_graphik4\\_pdf/wortgeschichten\\_ultima\\_mit\\_graphik4.pdf](http://germanistik.uni-mannheim.de/abteilungen/germanistische_linguistik/prof_dr_beate_henn_memmesheimer/publikationen_henn_memmesheimer/wortgeschichten_ultima_mit_graphik4_pdf/wortgeschichten_ultima_mit_graphik4.pdf) [Zuletzt geprüft am 28.03.2012].
- Histodillo (2012). *Armadillo: Historical Data Mining*. URL: <http://www.hrionline.ac.uk/armadillo/> [Zuletzt geprüft am 19.04.2012].
- Hockey, S. (1993). *Developing Effective Resources for Research on Texts: Collecting Texts, Tagging Texts, Cataloguing Texts, Using Texts, and Putting Texts in Context*. Literary and Linguistic Computing (8), Nr. 4, (S. 235–242).
- H-Soz-u-Kult (2012). *H-Soz-u-Kult / Kommunikation und Fachinformation für die Geschichtswissenschaften*. URL: <http://hsozkult.geschichte.hu-berlin.de/> [Zuletzt geprüft am 19.04.2012].
- Impact (2012). *Impact. Improving access to text : Home*. URL: <http://www.impact-project.eu/> [Zuletzt geprüft am 09.05.2012].
- Institut für Deutsche Sprache (2003). *COSMAS I - Korpusrecherche- und -analysesystem*. URL: [http://www.ids-mannheim.de/kl/projekte/cosmas\\_1/](http://www.ids-mannheim.de/kl/projekte/cosmas_1/) [Zuletzt geprüft am 11.01.2012].

- Institut für Deutsche Sprache (2010a). *COSMAS II - Das Korpusrecherchesystem des IDS*. URL:<http://www.ids-mannheim.de/cosmas2/projekt/service/flyer10.pdf> [Zuletzt geprüft am 11.01.2012].
- Institut für Deutsche Sprache (2010b). *COSMAS II - Übersicht*. URL:<http://www.ids-mannheim.de/cosmas2/uebersicht.html> [Zuletzt geprüft am 11.01.2012].
- Institut für deutsche Sprache und Linguistik der Humboldt Universität zu Berlin (2011). *Korpora — Institut für deutsche Sprache und Linguistik*. URL:[http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/links/korpora\\_links#franzoesisch](http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/links/korpora_links#franzoesisch) [Zuletzt geprüft am 02.05.2012].
- Jannidis, F. (2009). *TEI in a crystal ball*. *Literary and Linguistic Computing* (24), Nr. 3, (S. 253–265).
- Joint Information Systems Committee (2009). *Silchester Roman Town: A Virtual Research Environment for Archaeology*. URL:<http://www.jisc.ac.uk/whatwedo/programmes/vre1/silchester> [Zuletzt geprüft am 29.02.2012].
- Keim, I. (Hrsg.) (2002). *Soziale Welten und kommunikative Stile: Festschrift für Werner Kallmeyer zum 60. Geburtstag*, Studien zur deutschen Sprache (Band 22), Tübingen: Narr.
- Knorr-Cetina, K. (2002). *Wissenskulturen: ein Vergleich naturwissenschaftlicher Wissensformen*, Suhrkamp-Taschenbuch Wissenschaft (Band 1594), 1. Auflage, Frankfurt am Main: Suhrkamp. URL:<http://katalog.suub.uni-bremen.de/DB=1/LNG=DU/CMD?ACT=SRCHA&IKT=8000&TRM=84593485>\*.
- Küster, M. W.; Ludwig, C. und Aschenbrenner, A. (2009). *TextGrid: eScholarship und vernetzte Angebote TextGrid: eScholarship in a Network of Resources. It - Information Technology*. *it - Information Technology* (51), Nr. 4, (S. 183–190). URL:<http://www.oldenbourg-link.com/doi/pdfplus/10.1524/itit.2009.0540>.
- Leibniz-Zentrum für Psychologische Information und Dokumentation (2012). *Psychologie Information: Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID)*. URL: <http://www.zpid.de/> [Zuletzt geprüft am 02.05.2012].
- Liao, H.-t. und Petzold, T. (2010). *Analysing Geo-linguistic Dynamics of the World Wide Web: The Use of Cartograms and Network Analysis to Understand Linguistic Development in Wikipedia*. *Journal of Cultural Science* (3), Nr. 2, (S. 1–18). URL:<http://cultural-science.org/journal/index.php/culturalscience/article/view/44/128> [Zuletzt geprüft am 28.03.2012].
- Linten, M.; Woll, C. und Liebig, M. (2011). *Was sind die herausragenden Themen der beruflichen Bildung? Bibliometrische Analyse wissenschaftlicher Publikationen aus 2010 mit Hilfe der Literaturdatenbank Berufliche Bildung (LDBB)*. bwp@

- Berufs- und Wirtschaftspädagogik online, (S. 1–16).  
URL:[http://www.bwpat.de/10jahre\\_bwpat/linten\\_etal\\_10jahre-bwpat.pdf](http://www.bwpat.de/10jahre_bwpat/linten_etal_10jahre-bwpat.pdf) [Zuletzt geprüft am 31.01.2012].
- Lipp, A. (2009). *An integrated digital research environment: DFG perspectives*. Library Hi Tech (27), Nr. 4, (S. 496–504) [Zuletzt geprüft am 01.02.2012].
- Lucke, J. von und Geiger, C. (2010). *Open Government Data - Frei verfügbare Daten des öffentlichen Sektors*. URL:<http://www.zepelin-university.de/deutsch/lehrstuehle/ticc/TICC-101203-OpenGovernmentData-V1.pdf> [Zuletzt geprüft am 31.10.2011].
- Markdot (2012). *Gutenberg-Klammer und Copyright*. URL:  
<http://blog.markdot.com/tag/recht/> [Zuletzt geprüft am 09.05.2012].
- Meyer, T. (2011). *Virtuelle Forschungsumgebungen in der Geschichtswissenschaft*. Libreas, Nr. 18. URL:<http://edoc.hu-berlin.de/docviews/abstract.php?id=37869>.
- Miller, P. (31.10.2011). *Who is afraid of the GGG?*  
URL:[http://blogs.talis.com/nodalities/2007/11/who\\_is\\_afraid\\_of\\_the\\_ggg.php](http://blogs.talis.com/nodalities/2007/11/who_is_afraid_of_the_ggg.php)  
[Zuletzt geprüft am 31.10.2011].
- Mittler, E. (1994). *Auf dem Wege zu einem bibliothekarischen Gesamtkonzept: Neubau und historisches Gebäudeensemble der niedersächsischen Staats- und Universitätsbibliothek Göttingen*. In Poll, R. und Haller, B. (Hrsg.), *Bibliotheksbauten in der Praxis: Erfahrungen und Bewertungen*, (S. 223–241), Wiesbaden: Harrassowitz.
- Monumenta Germaniae Historica (2011). *Monumenta Germaniae Historica*. URL:  
<http://www.dmgh.de/> [Zuletzt geprüft am 22.08.2011].
- Nentwich, M. (2003). *Cyberscience: Forschung im Zeitalter des Internets*. Technikfolgenabschätzung (12), Nr. 3/4, (S. 73–78).  
URL:<http://www.itas.fzk.de/tatup/033/nent03a.pdf> [Zuletzt geprüft am 16.03.2012].
- Neuroth, H.; Aschenbrenner, A. und Lohmeier, F. (2007). *e-Humanities – eine virtuelle Forschungsumgebung für die Geistes-, Kultur- und Sozialwissenschaften*. BIBLIOTHEK Forschung und Praxis (31), Nr. 3, (S. 272–279).
- Neuroth, H.; Jannidis, F.; Rapp, A. und Lohmeier, F. (2009). *Virtuelle Forschungsumgebungen für e-Humanities. Maßnahmen zur optimalen Unterstützung von Forschungsprozessen in den Geisteswissenschaften*. BIBLIOTHEK Forschung und Praxis (33), Nr. 2, (S. 161–169).
- Nietzsche Source (2011). *Nietzsche Source*. URL: <http://www.nietzschesource.org/>  
[Zuletzt geprüft am 22.08.2011].
- O'Reilly, T. (09/30/2005). *What Is Web 2.0*.  
URL:<http://oreilly.com/pub/a/web2/archive/what-is-web-20.html?page=1>.

- Palmer, C. L. und Cragin, M. H. (2008). *Scholarship and disciplinary practices*. Annual Review of Information Science and Technology (42), Nr. 1, (S. 163–212). URL:<http://onlinelibrary.wiley.com/store/10.1002/aris.2008.1440420112/asset/1440420112ftp.pdf?v=1&t=gz9qui1r&s=54b2dfb23db0f1d1d829e4a3779ae7da41268bd9>.
- Palmer, C. L. und Neumann, L. (2002). *The Information Work of Interdisciplinary Humanities Scholars: Exploration and Translation*. The Library Quarterly (72), Nr. 1, (S. 85–117). URL:<http://www.jstor.org/stable/pdfplus/4309582.pdf> [Zuletzt geprüft am 05.03.2012].
- Palmer, C. L.; Tefteau, Lauren C. und Pirman, C. M. (2009). *Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development. Report commissioned by OCLC Research*. URL:<http://www.oclc.org/research/publications/library/2009/2009-02.pdf> [Zuletzt geprüft am 26.10.2011].
- Perseus Digital Library. URL: <http://www.perseus.tufts.edu/hopper/> [Zuletzt geprüft am 19.04.2012].
- Qualiservice (2012). *Qualiservice*. URL: <http://www.qualiservice.org/> [Zuletzt geprüft am 02.05.2012].
- Rapp, A. (2010). *TextGrid Virtuelle Forschungsumgebung in den Geisteswissenschaften*. URL:[http://www.werc.tu-darmstadt.de/fileadmin/user\\_upload/GROUP\\_WERC/2010\\_Tagung/TextGrid - Virtuelle Forschungsumgebung in den Geisteswissenschaften - Rapp.pdf](http://www.werc.tu-darmstadt.de/fileadmin/user_upload/GROUP_WERC/2010_Tagung/TextGrid_Virtuelle_Forschungsumgebung_in_den_Geisteswissenschaften_-_Rapp.pdf) [Zuletzt geprüft am 29.09.2011].
- Reagor, S. und Brown, W. S. (1978). *The application of advanced technology to scholarly communication in the humanities*. Computers and the Humanities (12), Nr. 3, (S. 237–246). URL:<http://www.springerlink.com/content/x433q28846125772/fulltext.pdf>.
- Reimer, T. und Carusi Annamaria (2010). *VRE Collaborative Landscape Study*. URL:<http://www.jisc.ac.uk/media/documents/publications/vrelandscape-report.pdf> [Zuletzt geprüft am 24.02.2012].
- Rieger, S. und Schöpflin, U. (2007). *"European Cultural Heritage Online" (ECHO) - eine Forschungsinfrastruktur für die Geisteswissenschaften*. Kunstchronik (60), Nr. 11, (S. 510–513). URL:<http://www.digizeitschriften.de/dms/gcs-wrapper>.
- Sahle, P. (2008). *eScience History (?)*. URL:[http://kups.ub.uni-koeln.de/2512/1/eScienceHistory\\_preprint.pdf](http://kups.ub.uni-koeln.de/2512/1/eScienceHistory_preprint.pdf) [Zuletzt geprüft am 28.09.2011].
- Schirnbacher, P. (2009). *Virtuelle Forschungsumgebungen - neue Formen der wissenschaftlichen Kollaboration. DGF/DINI-Workshop zur Förderung der wissenschaftlichen Informationslandschaft in Deutschland*.

- URL:<http://www.dini.de/fileadmin/workshops/dfg-dini-forschungsumgebungen-2009/schirnbacher.pdf> [Zuletzt geprüft am 29.09.2011].
- Schneider, S. und Kerchner, B. (2006). *Foucault: Diskursanalyse der Politik: eine Einführung*, 1. Auflage, Wiesbaden: Verlag für Sozialwissenschaften.
- Schreibman, S. (2009). *The Text Encoding Initiative. An Interchange Format Once Again*. URL:<http://computerphilologie.tu-darmstadt.de/jg08/schreibman.html> [Zuletzt geprüft am 15.11.2011].
- Schreibman, S. (2011). *A companion to digital humanities*, Blackwell companions to literature and culture (Band 26), [Nachdr.], Malden, Mass: Blackwell Publ. URL: <http://www.digitalhumanities.org/companion/>.
- Shadbolt, N.; Berners-Lee, T. und Hall, W. (2006). *The Semantic Web Revisited*. IEEE Intelligent Systems (21), Nr. 3, (S. 96–101). URL:<http://ieeexplore.ieee.org/ielx5/9670/34311/01637364.pdf?tp=&arnumber=1637364&isnumber=34311>.
- Short, H. (2005). *The Role of Humanities Computing: Experiences and Challenges*. Literary and Linguistic Computing (21), Nr. 1, (S. 15–27). URL:<http://llc.oxfordjournals.org/content/21/1/15.full.pdf#page=1&view=FitH>.
- Smith, D.; Rydberg-Cox, J. und Crane, G. (2000). *The Perseus Project: a digital library for the humanities*. Literary and Linguistic Computing (15), Nr. 1, (S. 15–25). URL:<http://llc.oxfordjournals.org/content/15/1/15.full.pdf#page=1&view=FitH>.
- Staats- und Universitätsbibliothek Bremen (2011). *Die Grenzboten. Digitalisierungsprojekt bewilligt*. URL: <http://www.suub.uni-bremen.de/ueberuns/neues-aus-der-suub/digitalisierungsprojekt-bewilligt/> [Zuletzt geprüft am 09.05.2012].
- Stone, S. (1982). *Humanities scholars: information needs and uses*. Journal of Documentation (38), Nr. 4, (S. 292–313) [Zuletzt geprüft am 05.03.2012].
- Taverna (2012). *Taverna - open source and domain independent Workflow Management System*. URL: <http://www.taverna.org.uk/> [Zuletzt geprüft am 29.03.2012].
- TextGrid (2012). *TextGrid*. URL: <http://www.textgrid.de/> [Zuletzt geprüft am 29.02.2012].
- TEXTvire (2012). *TEXTvire*. URL: <http://textvire.cerch.kcl.ac.uk/> [Zuletzt geprüft am 29.03.2012].
- Thaller, M. (1980). *Automation on Parnassus. CLIO - A Databank Oriented System for Historians*. Historical Social Research (15), (S. 40–65).
- The Linking Open Data cloud diagram (2011). *The Linking Open Data cloud diagram*. URL: <http://richard.cyganiak.de/2007/10/lod/> [Zuletzt geprüft am 31.10.2011].

- University of Southern Denmark (2012). *The Gutenberg parenthesis – print, book and cognition*.  
 URL:[http://www.sdu.dk/en/om\\_sdu/institutter\\_centre/ilkm/forskning/forskningsprojekte/gutenberg\\_projekt](http://www.sdu.dk/en/om_sdu/institutter_centre/ilkm/forskning/forskningsprojekte/gutenberg_projekt) [Zuletzt geprüft am 16.04.2012].
- Unsworth, J. (2000). *Scholarly primitives: What methods do humanities researchers have in common, and how might our tools reflect this?* Symposium on Humanities Computing: Formal Methods, Experimental Practice, May 13, King's College, London. URL:<http://people.lis.illinois.edu/~unsworth//Kings.5-00/primitives.html> [Zuletzt geprüft am 28.02.2012].
- Virtueller Raum Reichsrecht (2010). *Virtueller Raum Reichsrecht*. URL:  
<http://virr.mpg.de/virr/> [Zuletzt geprüft am 19.04.2012].
- W3C (2010a). *Semantic Web*. URL: <http://www.w3.org/standards/semanticweb/> [Zuletzt geprüft am 15.11.2011].
- W3C (2010b). *Semantic Web Case Studies and Use Cases*. URL:  
<http://www.w3.org/2001/sw/sweo/public/UseCases/> [Zuletzt geprüft am 06.03.2012].
- W3C (2012). *Linked Data*. URL: <http://www.w3.org/standards/semanticweb/data> [Zuletzt geprüft am 06.03.2012].
- Weller, K. (2009). *Ontologien: Stand und Entwicklung der Semantik für das World Wide Web*. Libreas, Nr. 15, 2, (S. 5–12). URL:<http://edoc.hu-berlin.de/libreas/15/weller-kathrin-5/PDF/weller.pdf> [Zuletzt geprüft am 22.03.2012].
- Wikipedia Generische Programmierung (17.04.2012). *Generische Programmierung*. URL:<http://de.wikipedia.org/w/index.php?oldid=99984209> [Zuletzt geprüft am 09.05.2012].
- Wills, G.; Gilbert, L.; Gee, Q.; Davis, H.; Miles-Board, T.; Millard, D.; Carr, L.; Hall, W. und Grange, S. *Towards grid services for a virtual research environment*. In Goodyear, P. (Hrsg.). (2005), *Fifth IEEE International Conference on Advanced Learning Technologies (ICALT '05)*, Los Alamitos, California, (S. 863–867). URL:<http://ieeexplore.ieee.org/ielx5/10084/32317/01508840.pdf?tp=&arnumber=1508840&isnumber=32317>.
- Wolstencroft, K.; Fisher, P.; De Roure, D. und Goble, C. (2009). *Scientific Workflows*. URL: <http://cnx.org/content/m32861/1.3/> [Zuletzt geprüft am 29.03.2012].
- Wörterbuchnetz (2011). *Wörterbuchnetz*. URL: <http://woerterbuchnetz.de/> [Zuletzt geprüft am 29.09.2011].
- Zeitschriftendatenbank (2012). *Zeitschriftendatenbank*. URL:  
<http://www.zeitschriftendatenbank.de/nc/aktuelles/einzelansicht/article/2012-04-11-5876/> [Zuletzt geprüft am 09.05.2012].

Zielinski, A.; Pempe, W.; Gietz, P.; Haase, M.; Funk, S. und Simon, C. (2009). *TEI documents in the grid*. *Literary and Linguistic Computing* (24), Nr. 3, (S. 267–279).  
URL:<http://llc.oxfordjournals.org/content/24/3/267.full.pdf#page=1&view=FitH>.



## **Eidesstattliche Erklärung**

Hiermit versichere ich, Julia Elisabeth Klein, geboren am 20.11.1979 in Trier, an Eides statt, dass ich die vorliegende Arbeit zur Erlangung des Titels des Biblilotheksassessors eigenständig und ausschließlich unter Zuhilfenahme der angegebenen Quellen und Hilfsmittel verfasst habe. Weiterhin versichere ich, dass die parallel eingereichten Print- und Digitalversionen dieser Arbeit inhaltlich vollständig identisch sind.

Unterschrift:

---

(Julia Elisabeth Klein)