Humboldt-Universität zu Berlin
Institut für Bibliotheks- und Informationswissenschaft

Berliner Handreichungen
zur Bibliotheks- und
Informationswissenschaft

Heft 290

Schöne Literatur binär kodiert

Die Veränderung des Text- und Dokumentbegriffs am Beispiel digitaler Belletristik und die neue Rolle von Bibliotheken

von
Sarah Dudek

Schöne Literatur binär kodiert

Die Veränderung des Text- und Dokumentbegriffs am Beispiel digitaler Belletristik und die neue Rolle von Bibliotheken

von
Sarah Dudek

Berliner Handreichungen zur
Bibliotheks- und Informationswissenschaft

Begründet von Peter Zahn
Herausgegeben von
Konrad Umlauf
Humboldt-Universität zu Berlin

Heft 290

Dudek, Sarah

Schöne Literatur binär kodiert : Die Veränderung des Text- und Dokumentbegriffs am Beispiel digitaler Belletristik und die neue Rolle von Bibliotheken / von Sarah Dudek. - Berlin : Institut für Bibliotheks- und Informationswissenschaft der Humboldt-Universität zu Berlin, 2011. – 84 S. : graph. Darst. - (Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft ; 290)

ISSN 14 38-76 62

Abstract:

Jahrhundertelang waren das Buch und der literarische Text untrennbar miteinander verbunden. Das beschriebene oder bedruckte Papier war Voraussetzung für den Begriff des klar abgegrenzten Dokuments. Im Bitstream, mit E-Books und digitalen Textkorpora, geht diese eindeutige Abgrenzung des Dokuments verloren. Die Übergangsphase vom analogen zum digitalen Paradigma ist von Imitationen des traditionellen Dokuments im Digitalen geprägt. In diesem Kontext ist digitale Belletristik – im Gegensatz zu naturwissenschaftlichen Publikationen – ein Bereich, der sich einem Neudenken im Digitalen besonders stark widersetzt.
Dabei eröffnen digitale literarische Texte vielfältige Möglichkeiten und bilden als Primärliteratur die Grundlage für neue wissenschaftliche Methoden und Disziplinen wie die Computerphilologie. Dass sich mit digitalem Text die Arbeitsweisen von Verlagen, Bibliotheken, Rechenzentren und Wissenschaften entscheidend verändern, ist bereits zu beobachten. In dieser Arbeit wird am Beispiel literarischer Texte dargestellt, wie digitale Umgebungen den traditionellen Dokumentbegriff auflösen, welche Auswirkungen die technologischen Veränderungen insbesondere auf die Rolle von wissenschaftlichen und öffentlichen Bibliotheken haben, und ein Ausblick gegeben, wie ein genuin digital gedachtes Dokument aussehen könnte.

Diese Veröffentlichung geht zurück auf eine Masterarbeit im postgradualen Fernstudiengang Master of Arts (Library and Information Science) an der Humboldt-Universität zu Berlin.

Online-Version: http://edoc.hu-berlin.de/series/berliner-handreichungen/2011-290

Inhalt

1 Einleitung 6
2 Ceci n'est pas un document – Thesen zu Text- und Dokumentbegriff unter digitalen Bedingungen 13
1. These zur Erosion des Dokumentbegriffs 16
2. These zum Wandel der Kulturtechniken 17
3. These zur Diskretheit 20
4. These zur Rolle des Textbegriffs 23
5. These zu technischen Formaten und Standards 29
6. These zur Rekonstruktion des Dokumentbegriffs 33
3 Digitale Belletristik als Primärliteratur für philologische Forschung in Wissenschaftlichen Bibliotheken 43
4 Digitale Belletristik als allgemeines Kulturgut in Öffentlichen Bibliotheken 60
5 Fazit 72
Literaturverzeichnis 78

1 Einleitung

Im Februar 2010 wurde in den Feuilletons ungewöhnlich emotional debattiert: Helene Hegemann habe in ihrem Debütroman „Axolotl Roadkill“ ganze Passagen von anderen Autorinnen und Autoren abgeschrieben, ohne diese Quellen kenntlich gemacht oder die Rechte eingeholt zu haben. Hegemann konterte in einer Pressemitteilung ihres Verlages: „Wenn da die komplette Zeit über reininterpretiert wird, dass das, was ich geschrieben habe, ein Stellvertreterroman für die Nullerjahre ist, muss auch anerkannt werden, dass der Entstehungsprozess mit diesem Jahrzehnt und den Vorgehensweisen dieses Jahrzehnts zu tun hat, also mit der Ablösung von diesem ganzen Urheberrechtsexzess durch das Recht zum Kopieren und zur Transformation.“ Ihre Verlegerin gab zudem zu bedenken, dass die 17-jährige Autorin „mit der 'sharing'-Kultur des Internets aufgewachsen“ sei.[1] Allerdings bleibt dabei unerwähnt, dass Hegemann ihren Roman eben nicht frei zugänglich – etwa unter einer Creative-Commons-Lizenz – im World Wide Web (WWW) veröffentlicht hat, sondern ihn als im traditionellen Sinne urheberrechtlich geschütztes Produkt von einem Verlag hat herausgeben und vermarkten lassen. Er wird in erster Linie als Printprodukt wirtschaftlich erfolgreich sein – zumal angesichts der Aufmerksamkeit, die ihm in der Presse zuteil wurde. Kurz vor der Verleihung des Preises der Leipziger Buchmesse, für den Hegemanns Roman nominiert war, veröffentlichte der Verband Deutscher Schriftsteller die „Leipziger Erklärung zum Schutz geistigen Eigentums“. Günter Grass (geb. 1927), Günter Kunert (geb. 1929), Sibylle Lewitscharoff (geb. 1954), Christa Wolf (geb. 1929) und andere warnten vor einer „fahrlässige[n] Akzeptanz von Rechtsverstößen im etablierten Literaturbetrieb“: Die Möglichkeiten neuer Medien, auch die des Internets änderten nichts an der Tatsache, dass der Schutz geistigen Eigentums nach wie vor uneingeschränkt Geltung und Priorität genieße. Und es heißt weiter: „Kopieren ohne Einwilligung und Nennung des geistigen Schöpfers wird in der jüngeren Generation, auch auf Grund von Unkenntnis über den Wert kreativer Leistungen, gelegentlich als Kavaliersdelikt

[1] http://www.boersenblatt.net/356651/ Die zweite Auflage des Romans wurde mit einem detaillierten Quellenverzeichnis versehen; die Rechte der verwendeten Passagen wurden nachträglich vom Verlag eingeholt. Eine Übersicht über die Debatte um Hegemanns Roman ist bei Perlentaucher zu finden: http://www.perlentaucher.de/blog/135_abgeschrieben_oder_eigenes_werk%3f_links_zum_streit_ueber_helene_hegemanns_roman_%27axolotl_roadkill%27

angesehen“.[2] Liegt also ein Generationenkonflikt zwischen „digital natives“ und „digital immigrants“ vor? Droht die „digitale Revolution“ womöglich sogar „die bürgerliche Kultur zu zerstören“, wie es in einem Leitartikel zum Thema in der *Welt* drastisch formuliert wird?[3]

Der Fall „Axolotl“ demonstriert die Verunsicherung, die in einer Zeit entsteht, in der sich Kulturtechniken und Konzepte unter digitalen Bedingungen wandeln, während die Masse an (belletristischer) Literatur, die in weiten Teilen der Gesellschaft rezipiert wird, den Publikationswegen der Gutenberg-Ära verhaftet bleibt. Der Begriff des Dokuments steht in dieser Print-Welt zentral und ist seit Jahrhunderten Garant für klar abgegrenzte und auch im juristischen Sinne festgeschriebenen Regeln folgende Entitäten in der gesellschaftlichen Kommunikation. Wie das französische Autorenkollektiv mit dem Pseudonym Roger T. Pédauque[4] in seinen Publikationen aus den Jahren 2003 bis 2007 betont, ist dieses Konzept des Dokuments unter digital-vernetzten Bedingungen radikalen Veränderungen ausgesetzt. Gradmann/Meister präzisieren diese These und beobachten, dass im digitalen Zeitalter das linear-zirkuläre Funktionsparadigma, beispielsweise bezogen auf den Publikationsprozess, zerfasert. Die Funktionsprägung durch traditionelle Kulturtechniken – Lesen und Schreiben – schwinde. Ein Erosionsprozess habe eingesetzt: in digital-vernetzter Umgebung werden die Grenzen des Dokuments undeutlich, woraus zwangsläufig auch eine Dekonstruktion des Dokumentbegriffs resultiert. Gleichzeitig findet jedoch Pédauque zufolge eine „redocumentarisation“ statt, „une novelle forme de documentarisation qui reflète ou tente de refléter une organisation post-moderne de notre rapport au monde“.[5] Dabei spiele das Dokument innerhalb der Veränderungen, die Pédauque in allen Bereichen der postmodernen Gesellschaft beobachtet, „un rôle clé, mais il s'est transformé au point que l'on peut se demander s'il s'agit encore de la même entité“

[2] https://vs.verdi.de/urheberrecht/aktuelles/leipziger-erklaerung/data/Leipziger-Erklaerung.pdf

[3] http://www.welt.de/die-welt/debatte/article6839014/Wenn-der-Autor-verschwindet.html

[4] RTP-DOC ist eine interdisziplinäre Forschergruppe, die in den Jahren 2003-2006 über ein „réseau thématique pluridisciplinaire“ des Centre national de la recherche scientifique (CNRS) organisiert war. Der Name ist ein Akronym: **R**éseau **T**hématique **P**luridisciplinaire „**Doc**ument et contenus : création, indexation, navigation“. Beteiligt waren über 100 WissenschaftlerInnen, u. a. aus den Disziplinen Informatik, Linguistik, Philosophie, Soziologie, Mathematik, Bibliotheks- und Informationswissenschaft. Zur Arbeitsweise von RTP-DOC vgl. den Aufsatz „Roger T. Pédauque au travail_“ von Jean-Michel Salaün (Pédauque 2006, S. 213-218), das Vorwort zu „La redocumentarisation du monde“ von Niels Windfeld Lund, den Prologue, das Ende der Einleitung und „S'inspirer de Roger T. Pédauque“ von Jean-Michel Salaün (Pédauque 2007, S. 7-14, S. 25, S. 17-23).

[5] Für eine Herangehensweise im Kontext des Umbruchs von skriptografischen zu typografischen Informationssystemen im 15. und 16. Jahrhundert, die eine ähnlich universale Perspektive einnimmt, vgl. Giesecke: „Medienwandel, Sinnenwandel, Kulturwandel und schließlich Sprachwandel gehen Hand in Hand“ (Giesecke, S. 13).

(Pédauque 2007, S. 16). Dass neben dem Dokumentbegriff auch der Textbegriff im Untertitel dieser Arbeit auftaucht, ist dieser Diagnose geschuldet und stellt bereits eine Hypothese dar: Der Textbegriff gewinnt gegenüber dem Dokumentbegriff an Bedeutung. Nichtsdestotrotz bietet der Dokumentbegriff in dieser hybriden Situation, in der analoge und digitale Kulturtechniken nebeneinander existieren und ineinandergreifen, einen Ausgangspunkt zur Analyse.[6]

Der Wert einer Beschäftigung mit dem Text- und Dokumentbegriff unter digitalen Bedingungen für Bibliotheken, Verlagswelt und (Geistes-)Wissenschaften liegt auf der Hand, denn in all diesen Bereichen spielt das Dokument eine zentrale Rolle. Für Bibliotheken sind das Organisieren, Zugänglichmachen und Verfügbarhalten von Dokumenten traditionell zentrale Aufgaben. Verlage treten als Produzenten dieser Dokumente auf, während für die Geisteswissenschaften Dokumente den Haupt-Untersuchungsgegenstand darstellen. Für all diese Bereiche verändert der Eintritt in die digitale Umgebung Arbeitstechniken und Selbstverständnis fundamental.

Der Sinn von Bibliotheken, zumal von physischen, wird seit Jahren in der gesellschaftlichen Diskussion infrage gestellt – auch wenn mit Blick auf spektakuläre neue Bibliotheksbauten wie der Malmö City Library (2000), der Seattle Public Library (2004), der Universitätsbibliothek Utrecht Uithof (2005), dem IKMZ Cottbus (2005), der Zentralbibliothek in Amsterdam (2007) oder dem Grimm-Zentrum in Berlin (2009) ein Umdenken zu bemerken ist. Angesichts der Probleme, die sich mit der Masse an Information ergeben, stehen Bibliotheken vor neuen Herausforderungen. Durch das Internet sind viele Dokumente schneller verfügbar geworden, der Link zum Volltext ist spätestens mit Google Books und Google Scholar fast schon zur Gewohnheit geworden.[7] Entwicklungen in Richtung Semantic Web beziehungsweise Linked Open Data zeugen von Bemühungen, in diesem Kontext neue semantische Zugangswege zu erschließen.[8] Für Nutzer jenseits der Wissenschaft stellen schon heute nicht mehr

[6] Natürlich sind auch andere Perspektiven denkbar, aus denen das Phänomen des Paradigmenwechsels durch digitale Medien betrachtet werden kann. Interessant wäre beispielsweise, die Markteinführung und gesellschaftliche Resonanz auf neue Endgeräte zu analysieren. E-Book-Reader werden totgesagt und zum Buch der Zukunft stilisiert, iPhone und iPad von Steve Jobs in religiös anmutenden Ritualen der Welt präsentiert. Diese Themenfelder können allerdings nicht in dieser Arbeit berücksichtigt werden.

[7] Andere Volltext-Suchmaschinen existieren zwar, sind jedoch weit weniger bekannt. Nach dem Ankauf von OAISTER durch OCLC und der Integration in die World-Cat-Suche wird sich das wohl ändern. Für den akademischen Bereich kann BASE (base.ub.uni-bielefeld.de) als weiteres Beispiel angeführt werden.

[8] Zur Unterscheidung der Termini „Semantic Web“ und „Linked (Open) Data“ gibt es unterschiedliche Ansichten. Es ist festzustellen, dass die Protagonisten des W3C in den letzten Jahren den Begriff „Semantic Web“ zu vermeiden und zunehmend durch „Linked Data“ zu ersetzen scheinen. Tim

klar abgegrenzte Dokumente, sondern miteinander vernetzte Informationseinheiten, die beliebig in unterschiedliche Umgebungen eingefügt werden können, die primäre Informationsquelle dar. Im WWW sind neue Formen des individuellen und kollektiven Schreibens, neue Formen der Nachnutzung von Inhalten wie Blogs, soziale Netzwerke und Wikis entstanden – das erfolgreichste Beispiel in diesem Kontext ist Wikipedia. Angesichts dieser Entwicklungen müssen zumindest den „digital natives“ traditionelle OPACs von Bibliotheken mit ihren Medienbrüchen und ihren bibliographischen Nachweisen, die in ihrem Aufbau die Überführung des Zettelkatalogs in eine digitale Umgebung erkennen lassen, längst als unverständliche und – im Privaten und wissenschaftlich vor allem im STM-Bereich – als überflüssige Relikte erscheinen. Natürlich haben sich auch die OPACs von Bibliotheken inzwischen verändert. Viele Bibliotheken bieten ihren Nutzern mittels Portalsoftware, Suchmaschinentechnologie, Linkresolvern und innovativen Oberflächen mit (national-)lizenzierten elektronischen Zeitschriften und E-Books vermehrt solche Volltexte an, die nicht frei zugänglich sind.[9] Für die Nutzer Wissenschaftlicher Bibliotheken ist dieser Service mithilfe von VPN auch als Fernzugriff möglich, Öffentliche Bibliotheken erlauben Fernzugriff meist über eine kurze Authentifizierungsanfrage, haben jedoch noch kein überzeugendes Äquivalent zu VPN gefunden – zumal sich bei ihrem nicht klar begrenzten Nutzerkreis andere rechtliche Probleme ergeben. Bei den Lizenzen handelt es sich zum größten Teil um wissenschaftliche Literatur, Belletristik dagegen kommt im digitalen Angebot Wissenschaftlicher Bibliotheken so gut wie nicht vor, in Öffentlichen Bibliotheken zumeist noch in nutzerunfreundlichen Formaten und einer wenig attraktiven Titelauswahl.

Die rechtliche Situation gestaltet sich in Bezug auf digitale belletristische Literatur schwierig: Findet die Open-Access-Bewegung für wissenschaftliche Texte inzwischen größeren Rückhalt, stellen sich Publikumsverlage, wie der Heidelberger Appell 2009 gezeigt hat, schützend vor ihre Autorinnen und Autoren und verteidigen ihren Status als wirtschaftliche Unternehmen. Die Folge sind Geschäftsmodelle und Portale wie Libreka oder die sogenannte „Onleihe“ mit nutzerunfreundlichen Formaten und rigidem Kopierschutz, die die Techniken der Printwelt imitieren. Im

Berners-Lee hat „Linked Data“ einmal als "the Semantic Web done right" bezeichnet (vgl. z. B. http://linkeddata.org/faq).

[9] Z. B. durch auf Lucene basierenden Produkten wie Primo und VuFind (letzteres ist ein Open-Source-Produkt) oder innovativen Oberflächen wie Bibscout, Touchpoint oder den Aquabrowser.

(literatur-)wissenschaftlichen Bereich wird mit zeitgenössischer Primärliteratur in digitalen Formaten bislang kaum gearbeitet. Allein die kleine Gemeinschaft der Computerphilologinnen und -philologen beschäftigt sich mit den neuen Möglichkeiten an Arbeitstechniken, die digitale Literatur bietet. Da die Computerphilologie durch die Editionswissenschaft beziehungsweise deren Beschäftigung mit großen Digitalisierungsprojekten geprägt ist, spielen jedoch – sicherlich auch aufgrund der rechtlich komplexen Situation – genuin digitale Texte und zeitgenössische Literatur bislang auch hier kaum eine Rolle. Dabei sind belletristische Texte als Primärliteratur die zentralen Quellen der Philologien und stellen die Bedingung für eine angemessene Rezeption der philologischen Forschungsliteratur dar. Ihr Status ist mit dem der Primärdaten in den Naturwissenschaften vergleichbar, die zunehmend zusammen mit den Forschungspublikationen – unter anderem durch die Förderung der Deutschen Forschungsgemeinschaft (DFG)[10] – digital zur Verfügung gestellt werden. Dass bislang kaum die Forderung laut wurde, die verwendete Primärliteratur in Verbindung mit der philologischen Forschungsliteratur digital verfügbar zu machen, ist wohl auch der noch vergleichsweise geringen Bedeutung des digitalen Publizierens in den traditionell monographien- und printzentrierten Geisteswissenschaften geschuldet. Trotzdem zeigen DFG-Großprojekte wie das VRE (Virtual Research Environment)[11] Textgrid – eine Virtuelle Forschungsumgebung, die durch Grid-Technologie leistungsfähige Strukturen für IT-gestütztes wissenschaftliches Arbeiten und Kooperation zwischen Forschergruppen bieten will –, dass auch für die Geisteswissenschaften an digitalen Infrastrukturen gearbeitet und ein Umdenken erwartet wird. Sollten sich digitale Arbeitstechniken in den Philologien zunehmend durchsetzen, stellt sich die Frage nach der langfristigen Verfügbarkeit und der Möglichkeit, dauerhaft auf diese Quellen referenzieren zu können. Offene technische Standards sind die Vorbedingung für die Ausbildung einer digitalen Geisteswissenschaft. Eine Untersuchung möglicher Dokumentmodelle, die die Grundlage für solche Standards bilden, sollte somit auch auf die Agenda der Philologien gelangen.

[10] Vgl. http://www.dfg.de/download/pdf/foerderung/programme/lis/positionspapier.pdf

[11] Zur Definition von VRE vgl. z. B. die aktuelle Landscape Study des JISC: „a VRE can be described as a set of web applications, online tools, systems and processes interoperating to facilitate or enhance the research process within and without institutional boundaries; it enables collaborative research activities beyond geographical barriers. Industry, universities, other research institutions and government are all involved in collaborative research projects“ (Carusi/Reimer, S. 12).

Das Themenfeld, das sich um den Dokument- und Textbegriff gruppiert, ist weit: Es reicht von rechtlichen Fragen über Format- und Enkodierungsaspekte und damit verbundenen Standardisierungsprozessen über die Bearbeitung komplexer Dokumentobjekte, die Referenzierung und Prozessierung von digitalen Dokumenten bis hin zu Themen wie Langzeitarchivierung und -verfügbarhaltung, Fragen zu geeigneten Oberflächen zum Zugang zu digitalen Dokumenten, Retrievalfunktionen und Endgeräten. Angesichts der Komplexität des damit nur grob abgesteckten Feldes soll in dieser Arbeit eine Eingrenzung getroffen werden. Gegenstand werden belletristische Texte, also Volltexte, und Metadaten nur insofern sein, als sie Teil digitaler Dokumente sind. Dabei werden (Retro-)Digitalisate, also digitale Dokumente, die – zum Beispiel mithilfe von OCR (Optical Character Recognition) – durch Digitalisierung eines analogen Mediums entstanden sind, weitgehend außen vor gelassen, weil genuin digitale Szenarien in den Blick genommen werden sollen. Dass dieser Fokus an manchen Stellen (auf Retrodigitalisate) auszuweiten ist, liegt an der noch schlechten Materialsituation in Bezug auf genuin digitale zeitgenössische Belletristik im wissenschaftlichen Bereich. Um den Text- und Dokumentbegriff zu untersuchen, wird weitestgehend auf Kodierung, mögliche Formate und damit verbundene Funktionsmodi eingegangen – im Mittelpunkt stehen XML-basierte Formate, aber auch semantisch basierte Technologien. Die daran angrenzenden rechtlichen, Archivierungs- und Retrievalaspekte sollen hierbei, soweit möglich, unberücksichtigt bleiben. Es wird nach der *differencia specifica* des digitalen im Vergleich zum analogen Dokument gesucht. Damit verbunden ist eine Reflektion über die Möglichkeiten, die digitale Dokumente/Texte bieten und die Frage, wie diese Potenziale realisiert werden. Inhaltlich geht es um digitale Literatur, die auf traditionellen Publikationswegen erscheint. Phänomene wie Netzliteratur, Handyromane und Ähnliches sollen schon deswegen nicht diskutiert werden, da sie bislang für Bibliotheken und Wissenschaft eine untergeordnete Rolle spielen und keine ähnlich breite Rezeption und keine ähnliche hohe Reputation besitzen wie Texte, die in Verlagen erscheinen.

Je nach Verwendungszweck der Literatur werden die Fragestellungen einerseits Wissenschaftliche Bibliotheken, Rechenzentren/Entwickler, (Publikums-)Verlage, Universitäten oder andere Forschungseinrichtungen betreffen, andererseits Öffentliche Bibliotheken, Entwickler und Publikumsverlage beziehungsweise ihre

Dachverbände. Da der Nutzen von digitaler Belletristik in den genannten Zusammenhängen ein anderer ist, spiegelt sich diese pragmatische Unterscheidung in der Struktur der Arbeit wider. Nachdem im ersten Kapitel auf der Basis des aktuellen informationswissenschaftlichen Forschungsstands Thesen zur Veränderung des Dokument- und Textbegriffs unter digitalen Bedingungen formuliert werden, fokussiert das zweite Kapitel auf digitale Belletristik als Primärliteratur für die wissenschaftliche Nutzung und die sich daraus ergebenden Herausforderungen für Wissenschaftliche Bibliotheken. Im dritten Kapitel werden demgegenüber digitale Belletristik als allgemeines Kulturgut und die Rolle von Öffentlichen Bibliotheken thematisiert.

2 *Ceci n'est pas un document* – Thesen zu Text- und Dokumentbegriff unter digitalen Bedingungen

Im Jahr 1929 entsteht René Magrittes berühmtes Gemälde „La Trahison des images“: die figürliche Darstellung einer Pfeife mit der kalligrafischen Bildunterschrift *Ceci n'est pas une pipe*.[12]

Nachdem Magritte in Michel Foucaults „Ordnung der Dinge“ die zentralen Begriffe der „Ähnlichkeit“ und „Gleichartigkeit“ aufgefallen waren, schreibt er Foucault 1966 einen Brief, in dem er auf ihre Verwandtschaft im Denken hinweist.[13] Foucault wird daraufhin zum Interpreten Magrittes und analysiert in seinem Essay „Ceci n'est pas une pipe“ aus dem Jahr 1973 die „Trennung zwischen figürlicher Darstellung (welche die Ähnlichkeit einschließt) und sprachlicher Referenz (welche die Ähnlichkeit ausschließt)“ (Foucault 1997, S. 25).

Die Kapitelüberschrift *Ceci n'est pas un document* sucht die Analogie zu Magrittes Verwirrspiel und Foucaults Analyse der komplexen semiotischen Beziehungen und Unterschiede zwischen Gegenstand, bildlicher Repräsentation und sprachlicher Referenz. Warum jedoch soll das Dokument, ein kulturelles Konzept, an die Stelle der Pfeife, eines Gegenstands, treten können? Die Theoriebildung zum Dokument geht für gewöhnlich von der Etymologie des Dokumentbegriffs, von seiner Herkunft vom lateinischen *documentum* beziehungsweise dessen Stamm *docere* (= lehren), aus (vgl. z. B. Lund 2004, S. 99; Pédauque 2006, S. 28), also von einer

[12] http://en.wikipedia.org/wiki/File:MagrittePipe.jpg#file
[13] Vgl. Walter Seitters Nachwort „Michel Foucault und die Malerei“ in Foucault 1997, S. 61-68.

pragmatischen Ebene, einer funktionalistischen Dokumentsicht: „documentation“ wird dementsprechend von Lund als Tätigkeit definiert, für die „some kind of instrument is needed“, an der Menschen beteiligt sind und in der eine „meaningful totality“ hergestellt wird (Lund 2004, S. 99ff.). Nichtsdestotrotz ist der Dokumentbegriff eng an das physische Medium gekoppelt: an das Papier (oder weiter gefasst: den Beschreibstoff), das es zu einer klar abgegrenzten Entität werden lässt und letztlich auch die pragmatische Ebene – seine Kreation und Distribution – bestimmt. Somit ist das traditionelle Dokument immer auch als Gegenstand zu begreifen. Unter digitalen Bedingungen wird das Dokument – ähnlich wie die Pfeife in Magrittes Gemälde – zunächst zur Repräsentation; es steht in einer mimetischen Beziehung zum traditionellen Dokument (vgl. McCarty, S. 29). Im Übergang vom analogen zum digitalen Zeitalter, so die zweite These in diesem Kapitel, wird das physische Dokument (als Bild oder in Formaten wie PDF) imitiert. Als bloße Repräsentation des Papierdokuments büßt es jedoch – und hier kommt die pragmatische Ebene wieder zum Tragen – seine Verlässlichkeit in einer Welt ein, die Kulturtechniken des Zugänglichmachens, des Organisierens und Archivierens von digitalen Dokumenten unter digitalen Bedingungen erst erprobt. Die Aussage *Ceci n'est pas un document* trifft in diesem Kontext durchaus zu: Es ist eben nicht mehr das Dokument, für das es sich ausgibt, sondern lediglich eine „trace iconique“ (Pédauque 2007, S. 17). Von „digitalen Dokumenten“ zu reden und Analogien zum traditionellen Dokument zu suchen, ist auf gewisse Weise – wenn es beim bloßen Ähnlichkeitsdenken bleibt – ein zirkuläres Unterfangen: ein Konzept aus der analogen Welt wurde – wie so viele – in die digitale übernommen (vgl. hierzu z. B. Puschmann), um nun herauszufinden, welche Analogien in Bezug auf die unterschiedlichen Elemente des Dokumentbegriffs im Digitalen bestehen. Deutlich wird dabei, dass der Begriff des digitalen Dokuments, unter anderem durch seine inflationäre Verwendung, kaum noch irgendeine Art von Konturschärfe besitzt. So beschwert sich bereits im Jahr 1996 ein Autor in dem amerikanischen Online-Magazin *Wired*:

> Have you noticed that the word *document* doesn't mean much these days? It covers everything from a text-only word processing file to a spreadsheet to a Java-soaked interactive Web page.
>
> It didn't used to be like this. A document was a piece of paper—such as a will or passport—with an official role in our legal system.
>
> But when the makers of word processors looked for something to call their special kind of files, they imported *document*. As multimedia entered what used to be text-only files, the word stretched to the point of meaninglessness. Just try to make sense of the file types Windows 95 puts into the Document menu entry. (Weinberger)

Es kommt zur Inflation und Erosion des Dokumentbegriffs im Digitalen. Im Zuge dieser Auflösung gewinnt zunächst der Textbegriff an Bedeutung, eben weil die Grenzen des Dokuments – zum Beispiel für Retrievalprozesse – durchlässig werden. Das bedeutet zugleich, dass die syntaktische (und eventuell auch die semantische) Ebene gegenüber der figürlich-materiellen Ebene zentral wird. Mit der Möglichkeit der Prozessierbarkeit von Daten und der Entwicklung von semantisch basierten Technologien stellen sich im Kontext der Künstlichen-Intelligenz-Forschung (KI) jedoch vermehrt Fragen nach den mit Dokumenten als menschlichen Artefakten, die in einer hermeneutischen Tradition stehen, verbundenen gesellschaftlich eingebetteten Funktionen: „S'il ne peut être 'vu' ou repéré, 'lu' ou compris, 'su' ou retenu, un document n'est d'aucune utilité“ (Pédauque 2007, S. 17). R. T. Pédauque loten – ausgehend von funktionalistisch-pragmatischen Dokumentdefinitionen – die Bedingungen einer Rekonstruktion des Dokumentbegriffs im Digitalen aus, die Möglichkeiten einer „redocumentarisation“ (Pédauque 2007), wobei ihre Vorstellungen wie auch die Ausführungen zu aktuellen Entwicklungen in dieser Arbeit notwendigerweise spekulativ bleiben müssen und sich der Gefahr aussetzen, schnell durch neuere Entwicklungen obsolet gemacht zu werden.

Um über den Dokumentbegriff nachzudenken, sind also verschiedene Annäherungen möglich: ein Phasenmodell des Wandels der Kulturtechniken (Pédauque 2006), eine Analyse der *differencia specifica* des digitalen im Vergleich zum analogen Dokument (Buckland), ein Rückgriff auf weite funktionalistisch ausgerichtete Dokumentbegriffe aus dem 20. Jahrhundert (Skare/Lund/Vårheim), eine Analyse mittels einer „méthode tridimensionnelle“, die auf unterschiedliche Dimensionen des Dokumentbegriffs – Form, Zeichen, Medium – gesondert eingeht (Pédauque 2006 und 2007) oder eine Fokussierung auf den Textbegriff (Renear). Damit sind bereits die drei großen Richtungen in der Theoriebildung der letzten Jahre benannt: die Schule um Niels W. Lund und die Dokumentationswissenschaft in Tromsø, die an Theorien von Paul Otlet und Suzanne Briet aus dem 20. Jahrhundert anschließt, die im Zuge der Text Encoding Initiative (TEI) entstandene Theoriebildung um Allen Renear (institutionell u. a. für eine gewisse Zeit an die Scholarly Technology Group an der Brown University gebunden, mittlerweile lehrt Renear an der Graduate School of Library and Information Science an der University of Illinois, Urbana-Champaign) und das französische Forschungsprojekt RTP-DOC, das zwischen 2003 und 2006 über ein

„réseau thématique pluridisciplinaire“ des Centre national de la recherche scientifique (CNRS) organisiert war, sowie Weiterentwicklungen der „pensées pédauquiennes“ durch Stefan Gradmann et al. (Gradmann/Meister).

Inspiriert von diesen unterschiedlichen Ansätzen, lassen sich folgende Thesen formulieren, die verschiedene Aspekte der Veränderung des Dokument- und Textbegriffs unter digitalen Bedingungen in den Fokus nehmen.

1. These zur Erosion des Dokumentbegriffs

Der Dokumentbegriff erodiert unter anderem, weil das physische Medium im Digitalen an Bedeutung verliert.

Das traditionelle Dokument ist ein Objekt. Theoretisch kann jedes Objekt – und sei es die Antilope im Zoo (Briet) – zum Dokument werden, wie Paul Otlet und Suzanne Briet Mitte des letzten Jahrhunderts dargestellt haben.[14] Diese weiten funktionalistischen Definitionen sind jedoch in Bezug auf digitale Textdokumente zu vernachlässigen. Das traditionelle Text-Dokument, der literarische Text, ist – handschriftlich oder gedruckt, als Schriftrolle oder Kodex – an das Medium Papier oder einen anderen Beschreibstoff gekoppelt. Seine Grenzen sind somit materiell klar abgegrenzt; es gibt monolithische Dokument-Einheiten. Das traditionelle Medium besteht aus einem Träger und einer Beschriftung, den geschriebenen oder gedruckten Zeichen (vgl. Pédauque 2006, S. 36).

Im Übergang vom analogen zum digitalen Signal verliert die physische Form ihre Definitionskraft für das Dokument: Im Digitalen ist alles ein Bitstream, jegliche Klarheit des Dokuments als physischer Form verschwindet. Fraglich ist, ob dieser Bitstream – in Analogie zum Papier – als Trägermedium aufgefasst werden kann (Gradmann 2009, S. 45), denn auch für Hardware, Datei oder Bildschirm als Medium ließe sich argumentieren (Pédauque, S. 38):

> It is, strictly speaking, sloppy terminology to speak of "digital media". There actually is no such thing as digital media, but only digital information. Digital information becomes "media" only by

[14] Paul Otlet, der belgische Begründer der Dokumentationswissenschaft, und die französische Bibliothekarin Suzanne Briet haben in den 30er bzw. 50er Jahren des 20. Jahrhunderts sehr weite Definitionen von Dokument geltend gemacht. Für Otlet ist Dokument all das, durch dessen Beobachten man informiert wird, das heißt, Dokumente können auch Skulpturen und Ähnliches sein (vgl. Otlet 1989). Für Briet ist ein Dokument Zeugnis eines Faktums, das heißt, jedes Zeichen, das intendiert ist, ein Phänomen zu repräsentieren, kann als Dokument gelten – Briets berühmtestes Beispiel ist das Beispiel der Antilope im Zoo: während eine wilde Antilope kein Dokument ist, kann eine Antilope im Zoo, als Studienobjekt im Kontext des Zeichensystems Zoo, als Dokument betrachtet werden (vgl. Briet 1951).

> the virtue of analog output; computer screens, loudspeakers, printers are analog output devices interfaced to the computer via digital-to-analog conversion hardware like video and sound cards or serial interfaces. (Cramer)

Statt Analogien aufzusuchen, soll hier vielmehr konstatiert werden, dass Signal und Medium im Digitalen untrennbar verbunden sind. Der Begriff des Mediums besitzt im Digitalen kaum mehr Aussagekraft und sollte daher vermieden werden. Dem Ansatz Pédauques folgend, dass der Übergang zum Digitalen nur einen Vektor gesamtgesellschaftlicher Prozesse darstellt, sehen Gradmann/Meister das Digitale „metonymically elevated to the status of one of the driving forces behind the change from a tangible goods oriented industrial society to a post-industrial society that deals in intangibles such as knowledge, information and services“ (Gradmann/Meister, S. 143).

Dementsprechend hat der Verlust des physischen Trägermediums auch Auswirkungen auf zentrale Kulturtechniken, die sich unter digitalen Bedingungen verändern. Der Leseprozess wird maschinenabhängig: War das Buch ohne weitere technische Hilfsmittel zugänglich, ist das Lesen digitaler Texte nur noch mittelbar über Technik möglich. Das Verhältnis von Original und Kopie wird bedeutungslos, sofern nicht die Zeit als zusätzliche Dimension mit hinzugenommen wird (Voß, S. 13). Die Stabilität des Dokuments ist im Digitalen nicht länger verbürgt. Archive, Universitäten und Bibliotheken stellen sich – zum Beispiel im Kontext der Gestaltung von Repositorien – Fragen nach Langzeitarchivierung und -verfügbarhaltung digitaler Dokumente, die bis heute weitestgehend ungelöst sind.

2. These zum Wandel der Kulturtechniken

Die zahlreichen Imitationen des analogen Mediums im Digitalen zeigen, dass wir uns derzeit noch in einer Übergangsphase vom analogen zum digitalen Paradigma mit den dazugehörigen Veränderungen der Kulturtechniken befinden.

Gradmann/Meister führen ein Phasenmodell ein, das die Evolution des Informationskontinuums im Übergang vom Druck- zum digitalen Zeitalter fassen soll. In diesem Modell werden drei Phasen unterschieden, wobei wir uns derzeit wahrscheinlich zwischen Phase zwei und drei befinden. Die erste Phase entspricht der Gutenberg-Galaxis, in dem das traditionelle (Papier-)Dokument zentral steht. Der Publikationsprozess, also Vorgänge der Kreation, Distribution und Rezeption von Dokumenten, verläuft linear und zirkulär um ein monolithisches gedrucktes

Informationsobjekt, das Dokument: Eine Autorin schreibt (Verfassen), das Manuskript wird begutachtet (Begutachten) ein Verlag druckt (Publizieren), die Bibliothek organisiert (Verwalten), ein Leser liest (Rezipieren), zum Zitieren wird geschrieben (Referenzieren), ebenso zum Annotieren (Annotieren). In der zweiten Phase, dem Emulationsmodus, befindet sich das digitale Dokument in einem druckanalogen Format (z. B. PDF). Der Publikationsprozess imitiert weitestgehend denjenigen des Gutenberg-Universums, auch wenn die einzelnen Arbeitsschritte in digitaler Umgebung stattfinden. Wehde zufolge sind damit die „Gutenbergsche und die digitale technologische Revolution strukturell vergleichbar“: „Auch Gutenberg suchte bei der formalen Entwicklung des Zeichensatzes des Buchdrucks, die formalen und ästhetischen Standards der handschriftlichen Buchstaben- und Textgestaltung zu imitieren bzw. zu vervollkommnen“ (Wehde, S. 7). Diese Phase des Wandels von Kulturtechniken ist offensichtlich generell vom Ähnlichkeitsdenken, vom Denken in Analogien geprägt. McCarty betont, dass die Analogie „inherently static“ ist: „it means either a type of relationship or an instance of one, never […] a process“ (McCarty, 29), womit zugleich die Problematik dieser Situation aufscheint. In der dritten Phase dominiert das XML-Dokument in vernetzter Umgebung 'born digital', wobei zu betonen ist, dass vieles in diesem Prozess noch nicht klar vorherzusagen ist. Der Publikationsprozess ist ein genuin digitaler, folgt aber den (in den Klammern angegebenen) tradierten Stadien: so wird der Inhalt in XML-Syntax mitsamt der zugehörigen Präsentation in XSL generiert (Verfassen), es finden – eventuell öffentliche – Prozesse elektronischen Annotierens statt (Begutachten), der Dokumentinhalt wird in verschiedenen Versionierungen mithilfe eines Identifikators stabilisiert (Publizieren), digitale Informationsobjekte werden in digitalen Bibliotheken organisiert (Verwalten), die Rezeption umfasst auch maschinelle Formen des Lesens, also des Prozessierens von Daten (Rezipieren), Zitieren findet möglicherweise nicht mehr als Replikation eines Dokumentteils, sondern in Form von Referenzieren auf Mikrostrukturen eines anderen Dokuments statt (Referenzieren), auch der Annotationsprozess ist digital, wobei unter anderem unklar ist, ob digitale Annotationen Teil des Dokuments werden (Annotieren). In dieser Phase ist das Informationskontinuum, der zirkulär-lineare Kreislauf noch immer intakt. Erste Auflösungserscheinungen sind jedoch schon zu beobachten: Die Reihenfolge der einzelnen Schritte ist im Digitalen keineswegs mehr in der beschriebenen Weise festgelegt. Im genuin WWW-basierten Kontinuum entsteht auch ein netzartiges

Funktionsmodell, in dem die Zwangsläufigkeit der linearen Abfolge verloren geht (vgl. Pédauque 2006; Gradmann/Meister). Ein Blick in die Zukunft lässt im Kontext von Linked Open Data eine weitere Auflösung dieses Funktionskontinuums erkennen, eine Radikalisierung der Erosion des Dokuments, ein Thema, das in den Ausführungen zu den folgenden Thesen noch aufgegriffen werden wird.

Letztlich entspricht heute jedem gedruckten Text ein genuin digitaler Text: Digitale Texte werden in der Herstellung von belletristischen Verlagen mithilfe von Layout-Programmen wie InDesign erstellt. Autorinnen und Autoren und Lektorat arbeiten in den meisten belletristischen Publikumsverlagen wiederum mit Textverarbeitungsprogrammen und zwar überwiegend in proprietären Formaten – ODF, das freie Open-Office-Format, hat sich bislang noch nicht durchsetzen können. Zwischen Herstellung, Lektorat und Autoren spielt PDF eine entscheidende Rolle, fixiert PDF doch den Text analog zu Druckfahnen und bildet somit eine Repräsentation des traditionellen Dokuments. Diese Arbeitsprozesse im Digitalen sind in der Welt der belletristischen Verlage lange Zeit nur auf die Herstellung des gedruckten Buches ausgerichtet gewesen, wobei Metadaten durchaus digital (z. B. mit Zwischenbuchhändlern oder der DNB) ausgetauscht werden. Während wissenschaftliche Verlage und Universitäten schon seit geraumer Zeit elektronisch publizieren, scheint den europäischen Publikumsverlagen erst in den letzten Jahren bewusst geworden zu sein, dass sie auch Produkte für den E-Book-Markt liefern müssen. Aus der Perspektive der Verlage war es dabei nur naheliegend, zunächst die stabilisierte digitale Druckvorlage – also in der Regel eine PDF-Datei – analog zu gedruckten Büchern zu vermarkten. Die Analogie zu den Funktionalitäten der Druckwelt wird durch Digital-Rights-Management-Modelle (wie dasjenige von Adobe) noch verstärkt, werden doch derart die technischen Möglichkeiten des digitalen Textes verhindert: in dieser Form steht der digitale Text den Möglichkeiten des analogen Textes, der kopiert werden kann, sogar noch nach, wobei Urheberrechtsfragen hier einmal außen vor gelassen werden sollen. PDF ist „antithetical to the spirit of the Web, being static rather than interactive, and difficult for machines to read, thus inhibiting the development of services that can link information between articles“ (Shotton 2009). Dass die digitalen Texte E-Books genannt werden und es nur innerhalb des jeweiligen Sprachspiels deutlich wird, ob damit der digitale Text oder die Hardware gemeint ist, ist ein Symptom des

Emulationsdenkens: Die Polysemie des Begriffs E-Book zeigt, dass Medium und Beschriftung früher eins waren und im Digitalen – schwer verständlich – auseinandergefallen sind.

XML-Systeme, die einen Workflow zwischen Lektorat, Herstellung und Vertrieb ermöglichen, werden – wie noch zu vertiefen sein wird – zumindest bei mittelgroßen bis kleineren Publikumsverlagen noch kaum genutzt. Genaueres hierzu wird in den nächsten Kapiteln noch thematisiert werden. Nicht unerheblich für die auch im wissenschaftlichen Bereich erkennbare Tendenz zur Simulation analoger Medien im Umgang mit belletristischen digitalen Texten waren große Digitalisierungsprojekte seit den 80er Jahren des 20. Jahrhunderts. Die Überführung eines analogen Textes in einen digitalen haben auch die Richtlinien der Text Encoding Initiative geprägt, in denen es nun zwar ausdrücklich heißt, dass sie auch auf Textmaterial anwendbar sind, das 'born digital' ist (vgl. Burnard/Bauman, S. Xxiii) – ob diese Kodierungsrichtlinien jedoch nicht schon aufgrund ihrer Herkunft dem Emulationsdenken verhaftet bleiben, ist im nächsten Kapitel genauer zu prüfen. Bislang sind unsere digitalen Bibliotheken jedenfalls noch mit „digital incunabula – digital objects whose form remains firmly rooted in traditions of print, with HTML and PDF largely mimicking the limitations of their print predecessors“ angefüllt (Crane).

3. These zur Diskretheit

Das digitale Dokument unterscheidet sich durch das Charakteristikum der Diskretheit grundlegend vom analogen Dokument und erfordert dennoch hermeneutische Tätigkeit.

Im Unterschied zu analogen Signalen, sind digitale Signale diskret, das heißt sie stellen eine lineare Abfolge von Einheiten – Bits – dar, die durch Leerstellen voneinander getrennt sind:

> In unübertreffbarer Zeichenökonomie taucht das Bit einzig als Gegensatz seiner eigenen Abwesenheit auf. Genau das nötigt indessen, wie schon die ersten Zahlenkolonnen in Leibniz' Handschrift zeigen, einen binären Code, längere Zeichenketten als alle anderen Symbolsysteme zu generieren. Endlos modulare Wiederholungen von Bits in der Zeit, bei seriellen Nachrichtenkanälen, oder auch im Raum, bei paralleler Datenverarbeitung, sind sein Betriebsgeheimnis. Wenn es, nach Lacan, das Zeichen von Zeichen überhaupt ist, ersetzt werden zu können, so maximiert ein binärer Code diese Austauschbarkeit der Teile bis zu dem (in Alphabeten ausgeschlossenen) Grad, daß jede Eins Ersatz der Null und jede Null Ersatz der Eins ist. (Kittler, S. 105f.)

Abgesehen von den erkenntnistheoretischen und metaphysischen Schwierigkeiten, die mit der Diskretheit einhergehen können (vgl. Gradmann/Meister, S. 143-145),

stellt der Übergang vom analogen zum digitalen Signal auch dadurch gerade für die Geisteswissenschaft einen grundlegenden Wandel dar, weil das semiotische System gewechselt wird: „le texte réduit à l'octet et l'image au pixel“ (Pédauque 2006, S. 13). Vom informatischen Standpunkt aus gesehen ist Text ein Datentyp und dem Umgang mit Text entspräche zum Beispiel die Speicherung und Prozessierung dieses Materials, also eines Bitstreams (vgl. Buzetti). Bezüglich des „Rezeptionsprozesses“ bedeutet das, dass Texte maschinenlesbar werden – wobei zu klären ist, was genau damit gemeint ist. Die „mediation of codification by mathematical programming [...] singles out digital text as an entirely novel form of written language“ (de Sousa, 235). Wichtig ist in diesem Kontext jedoch, zwischen Kodierungs- und Output-Ebene, zwischen informatischer und Rezeptionsebene zu unterscheiden (und somit den an Kulturtechniken und an menschliches Verstehen gekoppelten Dokumentbegriff mit in die Diskussion einzubringen): Schließlich ist zumindest für den Endnutzer die Repräsentation (von Text) und nicht dessen Kodierung zentral, das heißt, entscheidend ist nicht, ob Repräsentations- und Kodierungsebene eins ist wie im Fall des analogen Textes. Der Repräsentation von Text im alphabetischen Kode liegt eine alphanumerische Kodierung (heute zumeist auf Grundlage von Unicode als Zeichensatz) zugrunde – es gibt also letztlich eine doppelte Signifikantenebene, wobei die Dekodierung nur im Fall der Umwandlung des digitalen Kodes in die Repräsentation eines alphabetischen Kodes eindeutig verläuft, und ansonsten die hermeneutische Dimension, die komplexen Beziehungen zwischen Signifikant und Signifikat im Kontext unterschiedlicher Sprachspiele, bestehen bleiben. Es ist also zu betonen, dass die Betrachtung über Diskretheit auf Signalebene niemals die Signifikantenebene verlässt:

> „text“—if defined as a countable mass of alphabetical signifiers—remains a valid descriptor for both machine code sequences and human writing. In syntax and semantics however, machine code and human writing are not interchangeable. Computer algorithms are, like logical statements, a formal language and thus only a restrained subset of language as a whole (Cramer).

Der entscheidende Umbruch ist also nicht auf der epistemologischen Ebene angesiedelt, sondern betrifft die Veränderung der Kulturtechniken, vor allem des Lesens und des Organisierens von Information. Erst die Kodierung in diskreten Signalen ermöglicht Datenprozessierung. Und erst unter der Voraussetzung der Interoperabilität der Formate ist der Austausch von Daten, sind neue Distributionswege möglich. Während Daten in ihrer Rohform noch wenig nützlich

sind, kann Information – verstanden als nach bestimmten Mustern strukturierte Daten – nun auch von Maschinen nach logischen Regeln verarbeitet werden. Dadurch öffnen sich die Grenzen des Dokuments hin zur Kodierung, zum Text, zum Inhalt. Wenn sehr große Massen von Text mittels eines Algorithmus' durchdrungen werden können, ergeben sich neue Retrievalmöglichkeiten, die Möglichkeit bibliometrischer Analysen und computergestützter Textanalysen. Würde es sich dabei allein um maschinelle Verarbeitung von Daten handeln, wäre der Begriff des Dokuments sofort aufzugeben, da das Dokument nur im Kontext menschlicher Tätigkeit (Produktion, Rezeption) zu definieren ist. In Bezug auf bibliometrische und computerphilologische Analysen kann allerdings in der Tat davon gesprochen werden, dass sich eine neue Technik des Lesens entwickelt hat. Dabei bleibt auch hier die hermeneutische Ebene erhalten: erst eine von Menschen gestellte Forschungsfrage, die in einen Algorithmus übersetzt werden kann, liefert mittels Prozessierung großer Datenmengen Ergebnisse, die wiederum der menschlichen Analyse bedürfen. Anders formuliert: „as far as the humanities are concerned, all meaningful uses of computing are heuristic, and therefore […] mere 'delivery' is dangerously misleading“ (McCarty, S. 6). Die Potenziale computergestützter Analysen sind stark vom Granularitätsgrad, also von der internen Strukturiertheit und dem Auszeichnungsgrad der digitalen Dokumente abhängig, wobei die Auszeichnung, zumindest wenn es sich um semantische Auszeichnung handelt, nicht-automatisch stattfindet – auf diesen Zusammenhang wird im folgenden Kapitel noch eingegangen werden. Sowohl die Herstellung der Daten als auch das Programmieren der Algorithmen bedürfen also hermeneutischer Tätigkeit. Dies gilt auch für das Semantic Web, dessen Relationen (RDF, RDFS) und Ontologien[15] auf hermeneutischer Tätigkeit beruhen und dessen Software-Agenten

[15] Der aus der Philosophie stammende Begriff „Ontologie“ wurde bereits um 1980 von der KI-Forschung übernommen. Natürlichsprachliche Konzepte und Beziehungen zwischen diesen Konzepten werden in Ontologien organisiert: „Ontologien wurden im Umfeld der Künstlichen Intelligenz entwickelt und sind die zentralen Bausteine des Semantic Web: Mit ihnen kann Wissen einer Domäne formal repräsentiert und prinzipiell unabhängig von Programmen wieder verwendet werden. Sie beschreiben also Konzepte und ihre Beziehungen innerhalb einer Wissensdomäne und unterstützen Maschinen dabei, Inhalte im Web interpretieren zu können [...]. Sie bilden die höchste Stufe der semantischen Reichhaltigkeit [...] und setzen auf semantische Netze auf: Der Begriff „Ontologie ist aber keineswegs eindeutig und wird auf vielfältige Weise gebraucht, da unterschiedliche Zielsetzungen in einem Spektrum adressiert werden, das sich von maschineller Lesbarkeit von Daten (Automatisierung) bis hin zur Unterstützung von Menschen bei der Erfüllung komplexer wissensintensiver Arbeit (Wissensmanagement) erstreckt.“ (Blumauer/Pellegrini, S. 12) Eine Definition von „Ontologie“ im informatischen Sinne und einen Aufschluss über ihre Anwendungsfelder gibt z. B. Gruber: „Ontologies are part of the W3C standards stack for the Semantic Web, in which they are used to specify standard conceptual vocabularies in which to exchange data among systems, provide services for answering queries, publish reusable knowledge bases, and offer services to facilitate interoperability across multiple, heterogeneous systems and databases. The key role of on-

diesen Tätigkeiten erst nachgängig sind.

4. These zur Rolle des Textbegriffs

Der Textbegriff wird in digitaler Umgebung gegenüber dem Dokumentbegriff aufgrund der Auflösung der Dokumentgrenzen und der Trennung von Form und Inhalt wichtiger.

Während das Dokument, wie bereits gezeigt wurde, eng mit dem physischen Medium verbunden ist, das in digitaler Umgebung seine Bedeutung einbüßt, zielt der Textbegriff auf eine andere Ebene ab:

> According to the philological tradition, a document is first and foremost a material entity defined in opposition or complimentary to a work, or a text. The document is – in other words – the material manifestation of the text. [...] Another part of the historical tradition and the philological tradition emphasises the physical or material character of the document. Besides, we have the digital tradition that conceives of the document as a meaningful entity that does not depend on a physical support. (Ørom, S. 61)

Basiert das philologische Verständnis von Text – im Gegensatz beispielsweise zur technischen Sicht als einer diskreten Folge von Signalen – auf semiotischen Theorien, gleichgültig wie eng oder weit der Textbegriff jeweils gefasst ist,[16] wird das Dokument Ørom zufolge als Materialisierung von Text und als abgegrenzte komplexe semiotische Einheit verstanden. Wenn jedoch beispielsweise Algorithmen in einer Volltextsuche Dokumentgrenzen durchdringen oder Hyperlinks Dokumentgrenzen übertreten können, wird der materielle Inhalt des Dokuments, das heißt Daten (im technischen Verständnis), wird Text, das heißt die (narrativen) Strukturen (im semiotischen Sinne),[17] definitionskräftiger: „Le numérique a déplacé la question du support du document, qui en assurait la stabilité grâce à la fixité de l'inscription, vers la problématique des sa structure“ (Pédauque 2007, S. 17).

Der Eindruck, dass sich die Dokumentgrenzen auflösen, verstärkt sich in Bezug auf vernetzte Dokumente, wobei nicht-vernetzte Dokumente im Zuge der Etablierung des

---

tologies with respect to database systems is to specify a data modeling representation at a level of abstraction above specific database designs (logical or physical), so that data can be exported, translated, queried, and unified across independently developed systems and services. Successful applications to date include database interoperability, cross database search, and the integration of web services“ (Gruber).

[16] Vom sehr weiten Textbegriff im Sinne von Kristevas' Verständnis verschiedenster (auch nicht-sprachlicher) kultureller Systeme und Kodes als Text (vgl. z. B. Kristeva) über Positionen, die die Autonomie des literarischen Textes zu bewahren suchen und Intertextualität als sich materialisierende Bezüge zwischen Texten als sprachlich-fixierten Entitäten untersuchen (vgl. z. B. Genette), bis zur Eingrenzung des Begriffs, die „Textualität an die gesellschaftliche Aufbewahrung von Wissen“ bindet (Knobloch, S. 46).

[17] Vgl. z. B. Pédauques Definition von „contenu“ als „le codage alphabétique du discours et sa structuration“ (Pédauque 2006, S. 91).

WWW im letzten Jahrzehnt weitgehend an Bedeutung verloren haben. Im Kontext von (X)HTML-Seiten im WWW wird meist auf den Begriff des Dokuments verzichtet und derjenige der Ressource verwendet, die das World Wide Web Consortium (W3C)[18] schlicht als „anything that can have an identifier“ definiert (Booth et al. 2004). Im WWW verweisen Hyperlinks auf andere Ressourcen und bringen denjenigen, der ihnen folgt, – um in den im Kontext des WWW üblichen räumlichen Metaphern zu sprechen – an einen anderen Ort (im Sinne der URL). Trotz der Euphorie, die in den 1990er Jahren im Zuge früher HTML-Enkodierungsprojekte zu spüren war, die in Gestalt des Hypertextes poststrukturalistische Theorien von Barthes, Foucault, Derrida und Bachtin im Sinne von textueller Offenheit, Nichtlinearität und Intertextualität verwirklicht sahen, wurden diese Erwartungen kaum eingelöst: HTML-Archive haben sich als nicht besonders flexibel erwiesen; vielmehr werden durch Hyperlinks Leserichtungen aufgezwungen (vgl. Schreibman 2002, S. 285). Ted Nelson, der Erfinder von Hypertext, kritisiert die heutige – verarmte – Form von Hyperlinks im WWW, die stets nur in *eine* Richtung weisen und an einen virtuellen Raum und nicht an ein Dokument gebunden sind, was häufig den berüchtigten „Fehler 404“ zur Folge hat, der auftritt, wenn ein Dokument nicht mehr auf dem angefragten Server liegt (vgl. beispielsweise Nelson 2007). Fallen in der URL tatsächlich noch die Funktionen des Identifizierens und Lokalisierens zusammen, wird diesem Problem inzwischen mit DOIs (Digital Object Identifier) und URNs (Uniform Resource Name), also Identifiern ohne Lokalisierungsfunktion, die Resolver-Dienste übernehmen, begegnet.[19] Nichtsdestotrotz stellen Hyperlinks im Vergleich zum Zitat, das fremden Text ins Dokument integriert, zur bibliographischen Angabe, die einen Verweis auf ein anderes Medium darstellt, und zur Fußnote, die intratextuelle Sprünge dirigiert, etwas ganz und gar Andersartiges dar. Sie haben nicht nur die Funktion zu verweisen, sondern sind Ausgangspunkt für Aktion: Sie öffnen das Dokument, sie verlagern das Lesen an einen anderen Ort, ohne dass in jedem Fall deutlich wird, was nun als in den Text integriert gelesen werden sollte und was lediglich als Beleg gedacht ist. Wenn also die Grenzen des Dokuments durch

[18] Das W3C wurde 1994 gegründet und besteht aus Interessenvertretern aus Industrie und Wissenschaft. Es entwickelt und empfiehlt Standards für das WWW, u. a. für Sprachen und Technologien.

[19] Dass der „Fehler 404“ als notwendiges Übel im Zuge des dezentralen Wachstums des WWW in Kauf genommen wurde, klingt beispielsweise im folgenden Zitat an: „Decentralization requires compromises: the Web had to throw away the ideal of total consistency of all of its interconnections, ushering in the infamous message 'Error 404: Not Found' but allowing unchecked exponential growth“ (Berners-Lee et al., S. 37).

Hyperlinks sowie die bereits thematisierte Möglichkeit der Prozessierung des Dokumentinhalts porös werden, tritt der Inhalt – verstanden nicht als semantischer Inhalt, sondern zunächst als Daten –, in den Vordergrund.

Auf der Textebene selbst zerfallen Form und Inhalt durch die Trennung von strukturierten Daten und Layout, denn in digitalen Dokumenten lässt sich prinzipiell die Datenhaltung von der Präsentation der Daten, strukturelles von typografischem Markup trennen. Fügt HTML als Auszeichnungssprache dem Dokument oder dem darzustellenden Inhalt Informationen zur deren Darstellung im Browser zu, zielt XML ganz auf die logische Struktur von Dokumenten ab. Im XML-Dokument werden zwei Ebenen unterschieden: die Ebene der logischen Struktur in XML und die formale Repräsentation, das Stylesheet, zum Beispiel in XSL(-FO) oder CSS. Der Einfluss von XML als Syntax für die Modellierung strukturierter Dokumente und XHMTL als XML-Sprache nimmt durch die Standardisierungbestrebungen des W3C in diesem Bereich kontinuierlich zu. Ein entscheidender Unterschied zwischen XML-Dokumenten und HTML-Dokumenten besteht darin, dass zwar beide für Menschen lesbar, jedoch nur XML-Dokumente für Maschinen prozessierbar sind. Somit ist X(HT)ML derzeit das tragende Fundament von Linked Data.[20]

Dass XML-Dokumente neben dem Austausch von Daten auch für Cross-Publishing-Zwecke besonders geeignet sind, liegt an der bereits erwähnten Trennung von Repräsentation und Inhalt. Anhand der Auszeichnung der logischen (nicht der grafischen) Struktur des XML-Dokuments kann ein im XML-Dokument zu benennendes Stylesheet die ausgezeichneten Strukturen grafisch umsetzen, wodurch viele unterschiedliche Layout- und Ausgabevarianten eines Inhalts einfach umzusetzen sind. Eine solche Trennung des Layouts vom Inhalt resultiert aus Textmodellen wie dem OHCO-Modell (Ordered Hierarchy of Content Objects), das vor allem von Renear et al. propagiert wird und auf dem auch die Richtlinien der Text Encoding Initiative und ihrer auf XML basierenden Auszeichnungssprache für Texte basieren:

> The essential parts of any document form what we call „content objects," and are of many types, such as paragraphs, quotations, emphatic phrases, and attributions. Each type of content object usually has its own appearance when a document is printed or displayed, but that appearance is superficial and transient rather than essential—it is the content elements themselves, along with their content, which form the essence of a document. (DeRose et al., S. 3)

[20] Linked Data wäre auch ohne XML, lediglich auf RDF basierend vorstellbar, wird sich in dieser Variante jedoch voraussichtlich nicht durchsetzen.

Ein solches Verständnis von Text berücksichtigt jedoch in keiner Weise die Materialität beziehungsweise die Signifikantenebene von Zeichen und Zeichenbeziehungen, die sich insbesondere in literarischen Texten durch die grafische Darstellung herstellen und deren Bedeutung sich unter anderem an synoptischen digitalen Editionen zeigt, die eine Zusammenschau von transkribiertem Text und Faksimiles von Manuskripten oder gedruckten Ausgaben anbieten. Es ignoriert, dass „Schriftcharaktere selbst als semantische Größen wirken und die Anordnung der Schriftzeichen die Darstellung semantischer Abläufe, Wertigkeiten und Beziehungen zu leisten vermag“ (Wehde, S. 11).

Der Eindruck der Dominanz des Textes beziehungsweise des Inhalts gegenüber dem Dokument wird mit Blick auf semantisch basierte Technologien verstärkt. Das W3C will das „web of documents“ langfristig durch ein „web of data“ ablösen beziehungsweise dem „web of documents“ das „web of data“ hinzufügen (vgl. z. B. W3C 2010). Ziel ist es, in verstärktem Maße Daten im WWW so aufzubereiten, dass sie von Computern verwertet werden können: „To date, the Web has developed most rapidly as a medium of documents for people rather than for data and information that can be processed automatically. The Semantic Web aims to make up for this“ (Berners-Lee et al., S. 37). Zu den basalen Technologien zählen in diesem Kontext neben XML das Resource Description Framework (RDF) zur Beschreibung für WWW-Ressourcen und RDFS (RDF Schema), das als Grammatik von RDF fungiert und bereits einfache Hierarchiebildungen und Ähnliches ermöglicht. RDF besitzt eine Tripel-Struktur, wobei jedes Tripel der Struktur von Subjekt, Objekt und Prädikat ('property') folgt (vgl. Klyne/Carroll 2004). Subjekt, Objekt und Prädikat sind durch eine URI klar definiert (das Objekt kann auch ein Literal sein). Durch diese Struktur lassen sich vielfältige Beziehungen zwischen Ressourcen im WWW herstellen, die meist in einer Graphen-Struktur abgebildet werden. Ontologiesprachen wie OWL (Web Ontology Language) stellen darauf basierend komplexere Wissensrepräsentations-Systeme her. Bereits bestehende Dokumentationssprachen lassen sich mittels SKOS (Simple Knowledge Organisation System) relativ leicht ins Semantic Web überführen und nachnutzen. Auffällig ist in diesem Kontext, dass von informatischer Seite Dokumente mit einer Signifikantenebene identifiziert werden, von der sie das „web of data“ befreit zu sehen scheinen: „Whilst HTML provides a means to structure and link documents on the Web, RDF provides a generic, graph-

based data model with which to structure and link data that describes things in the world“ (Bizer et al., S. 37); „[o]n the Semantic Web, URIs identify not just Web documents, but also real-world objects like people and cars, and even abstract ideas and non-existing things like a mythical unicorn. We call these *real-world objects* or *things*“ (Sauermann/Cyganiak). Dass durch Linked Data statt Dokumenten endlich die „wirklichen“ Dinge in der Welt beziehungsweise die Worte, die auf die Dinge der Welt zeigen, und vermeintlich universelle Konzepte, die an (neo-)platonische Vorstellungen oder auch an Leibniz' Monaden erinnern, verbunden werden können, basiert auf einer „(technically) high-level nominalist regression“ (Gradmann/Meister, S. 147), die jegliches semiotische Modell ignoriert und vor allem Text als eigenständiges semiotisches System unberücksichtigt lässt. Problematisch sind dabei nicht in erster Linie die – beispielsweise mithilfe von bibliothekarischen Metadaten oder Social-web-Technologien entstehenden – Ontologien, sondern das „ontology-mapping“ beziehungsweise die „data fusion“, also „the process of integrating multiple data items representing the same real-world object into a single, consistent, and clean representation“ (Bizer et al.), das Van Harmelen zu Recht als „the Achilles Heel of the Semantic Web“ (Van Harmelen) bezeichnet. Beim Mapping werden nicht nur Daten miteinander verbunden (was wäre das Kriterium dafür?), sondern Konzepte (vgl. Pédauque 2006; Pédauque 2007). Das informatische Sprachkonzept, das David Golumbia als „Computationalismus“ bezeichnet (vgl. Golumbia 2009), überträgt die Funktionsweisen von Computern auf die natürliche Sprache. Die Vorstellung von festen denotativen Bedeutungsbeziehungen, das Ignorieren der Kontextabhängigkeit von Sprache (im Sinne Wittgensteinscher Sprachspiele) – von Phänomenen wie Polysemien und Ausdrucksmitteln wie Ironie –, von Rezeptionsprozessen (vgl. beispielsweise die Theorien der Konstanzer Schule), der Unübersetzbarkeit kultureller Konzepte, der Problematik abgeschlossener Semantiken, fester Klassifikationen, statischer „Ontologien“ (vgl. die Traditionslinie von Nietzsches Sprachkritik bis zur Frankfurter Schule) lassen diesen Ansatz verkürzt und problematisch erscheinen:[21]

> Bedeutungen kommen nicht den Dingen von außen wie eine Aufschrift hinzu, sondern sie erwachsen sozusagen aus unserem Umgang mit ihnen in der Welt als ein „Bezugsgewebe“

[21] „Es ist faktisch unmöglich, dieses System [dasjenige der kulturellen Einheiten] zu definieren, zu beschreiben und zu erschöpfen, und zwar nicht nur wegen seines Umfangs, sondern auch deshalb, weil im Kreis der unbegrenzten Semiose die kulturellen Einheiten sich entweder aufgrund neuer Wahrnehmungen oder durch entstehende Widersprüche in ihren Korrelationen beständig neu strukturieren (darin besteht das Leben der Kultur)“ (Eco, S. 179f.).

> zwischen Handelnden und Sprechenden [...]. Die „Bedeutung von etwas“ ist immer aus einer Bedeutungsganzheit zu verstehen, in der sie eingebettet ist und die wir miteinander teilen. Die menschliche Welt ist eine solche „bedeutsame“ Welt, in der sowohl die gegenständliche „Dingwelt“ des von uns künstliche Hergestellten als auch die uns gegebene Naturwelt mit eingewoben in unserem Reden und Handeln sind (Capurro, S. 529).

In einer Polemik Nelsons liest sich dieser Zusammenhang folgendermaßen:

> today's nightmarish new world is controlled by "webmasters", tekkies unlikely to understand the niceties of text issues and preoccupied with the Web's exploding alphabet soup of embedded formats. [...] And the "semantic web" means that tekkie committees will decide the world's true concepts for once and for all. (Nelson o. J.)

Mit der Konzentration auf kleinere Dateneinheiten im Kontext von Linked Open Data und ihrer Repräsentation in Ontologien tritt das Dokument zunehmend hinter seinem Inhalt zurück. XML-Dokumente werden mittels semantisch basierter Technologien wie den Gleaning Resource Descriptions from Dialects of Languages (GRDDL) durchdrungen und zerstückelt: Daten, die in RDF abgebildet werden können (z. B. Namensräume), werden automatisch aus einem Dokument extrahiert (vgl. Connolly). Fraglich ist bislang, wie das WWW konsequent in Linked Data aufgelöst werden kann – denn an eine automatische Überführung von XML-Dokumenten in RDF-Statements ist nur zu denken, wenn eine Interoperabilität durch eine identische Verwendung von Elementen in einer Masse von XML-Dokumenten sichergestellt ist. [22] Zudem ist bislang nicht abzusehen, wie sich in sehr großen Triplestores Performance und Skalierbarkeit gestalten.

Auch im bibliothekarischen Bereich zeigt sich das Umdenken vom Dokument zum Inhalt im digitalen Paradigma. Im Gegensatz zu den „Regeln für die alphabetische Katalogisierung“ (RAK) basiert das neue internationale Regelwerk „Resource Description and Access“ (RDA) auf dem theoretischen Modell der „Functional Requirements for Bibliographic Records“ (FRBR). Die FRBR haben einen funktionalistischen Ausgangspunkt; entscheidend ist die Art und Weise, wie bibliographische Angaben genutzt werden. Daher wandert der Fokus von der Titelaufnahme, die auf ein Dokument verweist, zum Inhalt, zu „the 'things' the data describe“ (IFLA Study Group on the Functional Requirements of Bibliographical Records, S. 9). Die RDA werden gegenüber ihrem Vorläufermodell, der zweiten Ausgabe der Anglo-American Cataloguing Rules (AACR2), explizit als „designed for the digital world“ (Joint Steering Committee for Development of RDA) herausgestellt – wahrscheinlich wäre zutreffender: für eine hybride Welt. Waren früher das Werk

[22] Einen Schritt in diese Richtung geht Calais (www.opencalais.com): so wandelt der Calais Web Service z. B. unstrukturierte Dokumente in RDF um.

und die davon verfügbaren Exemplare, also die Dokumente, zentral, wird nun der Ausgang beim Informationsobjekt genommen: es wird zwischen Inhalt und seiner Darstellung unterschieden, die sich beispielsweise darin zeigt, dass an ein Informationsobjekt mehrere Expressionen (z. B. eine französische Originalfassung und eine deutsche Übersetzung und eine überarbeitete französische Fassung aus einem späteren Jahr) und Manifestationen (z. B. eine Buchausgabe, eine PDF-Ausgabe, eine HTML-Ausgabe) gebunden werden können.[23]

Unter digitalen Bedingungen verliert das Dokument als abgegrenzte Einheit an Bedeutung; sein Inhalt wird zum Zentrum der Aufmerksamkeit, wobei noch ungeklärt ist, wie der Inhalt, wie der Text, der nun ohne klare Grenzen in vernetzter Umgebung vorliegt, die pragmatisch-gesellschaftlichen Funktionen des Dokuments übernehmen kann – ein Grund, aus dem Pédauque einen Prozess der „redocumentarisation" fordern, den die sechste These zum Thema machen wird.

5. These zu technischen Formaten und Standards

Technische Formate und Standards basieren auf Text- und Dokumentmodellen, die Auswirkungen auf den möglichen Umgang mit digitalen Dokumenten haben, wobei die reine Binärkodierung von Texten einen simplifizierenden Textbegriff befördert, der durch Markup in Analogie zum überlieferten Textbegriff gebracht wird.

> [T]he computer as tool and medium is not neutral, but rather informs […] the work that one does with it, if only by setting possibilities and limits on what can be done (or even thought). It calls for critical awareness. Like the historians of computing, digital scholars must learn to read software to elicit the history and practice that it embodies. (Mahoney, S. 132)

Die Aussage Mahoneys lässt sich noch verstärken: Informationstechnologie gibt nicht allein Möglichkeiten und Grenzen vor, sondern ist selbst historisch entstanden und bildet vorherrschende Diskurse und Denkweisen ab, die in der Folge wiederum den Umgang mit Information, Text und Dokumenten bestimmen. Pédauque betonen daher, dass „il est nécessaire de quitter toute conception neutraliste de la technique" (Pédauque 2006, S. 96). Aus dieser Perspektive lassen sich beispielsweise die hierarchischen Modelle, die die Informationstechnologie in fast allen Bereichen bestimmen – angefangen von der Organisationsstruktur von Dateien über den Aufbau von Markup-Metasprachen bis zur Architektur des WWW –, nicht als technische Vorgabe sehen, sondern vielmehr als Denkweise, die zumindest in der

[23] Ob Bibliothekskataloge sich über FRBR und RDA dem Bereich Linked Open Data annähern können, um vielleicht eines Tages als Informationsobjekte darin aufzugehen, ist indessen fraglich (vgl. Gradmann 2005).

westlichen Welt eine lange Tradition besitzt. Für Nelson stellt die „simulation of hierarchy“ die „principal tradition in the computer world“ dar (Nelson 2007). McCarty bezeichnet die hierarchische Ordnung als „layer-structure“, die uns zum Beispiel „physical objects in hierarchical groups roughly by size, from galaxies to subatomic particles, each a component of the next“ sehen lässt und zur Folge hat, dass wir „the biological, social and psychological worlds similarly, by species, classes and so forth“ organisieren (McCarty, S. 195). Auch die traditionellen bibliothekarischen Klassifikationssysteme sind ein besonders gutes Beispiel für diese Denkweise. Auf der Ebene der Markup-Sprachen zeigt sich dieses Denken vor allem im Aufbau von XML, einer Sprache, die stets „wohlgeformt“ zu sein hat, das heißt zum Beispiel, dass alle Elemente (Tags) geschlossen werden und sauber geschachtelt sein müssen, es darf keine Überlappungen und nur ein Wurzelelement geben. Die Wohlgeformtheit eines XML-Dokuments wird von XML-Parsern, von Programmen, die ein XML-Dokument einlesen, überprüfen und verarbeiten, überwacht. Für Nelson ist XML „not an improvement but a hierarchy hamburger. Everything, everything must be forced into hierarchical templates!“ (Nelson o. J.)

Auch den verschiedenen Formaten liegen Daten- und Textmodelle zugrunde, wobei „nicht das konkrete Format, sondern das dahinterstehende Daten- bzw. Informationsmodell“ entscheidend ist (Voß). Renear trifft hinsichtlich unterschiedlichen Textmodellen folgende Unterscheidung:

> one could model text as a sequence of graphic characters, as in the "plain vanilla ASCII" approach of Project Gutenberg; as a combination of procedual coding and graphic characters, as in a word processing file; as a complex of geometric shapes, as in "vector graphics" format of an image of a page on which the text is written; as a pure image, as in a raster image format (JPEG, GIF, etc.); or in a number of other ways. (Renear 2004, S. 225)

Wenn einmal von den Bildrepräsentationen von Text, die aufgrund ihrer eingeschränkten Funktionalitäten zu vernachlässigen sind, abgesehen wird, ist Text vom informatischen Standpunkt her zunächst ein Datentyp, eine Folge von alphanumerischen Zeichen.[24] Wie ist diese Sicht auf Text jedoch mit der philologisch fundierten Auffassung zu vereinen, dass ein Text eben „not an undifferentiated sequence of words, much less of bytes“ (Burnard/Bauman, S. xxxiii) darstellt? Das Datenmodell von Text ist mit literarturwissenschaftlichen Textmodellen in dieser Form nicht kompatibel. Buzetti zufolge ist die Struktur der digitalen Repräsentation

[24] Daher trifft die Aussage, dass „Schrift und Bild [...] für die technologische Be- bzw. Verarbeitung in der EDV gleiches Ausgangsmaterial“ darstellen nur insofern zu, dass sie beide „denselben – bisher ungeahnten – visuellen Manipulationen unterworfen werden“ (Wehde, S. 9). Auf Kodierungsebene sind Text (als alphanumerische Zeichenkette) und Bild grundsätzlich unterschiedliche Datentypen.

entscheidend, wenn es um die Einschätzung adäquater Bedingungen geht, weil erst sie die automatische Prozessierung des Informationsgehalts des Dokuments ermöglicht:

> The representation's form must serve the analytical operations necessary to the study of the text. […] An adequate digital text representation must therefore be compatible with the application of the formal procedures of information processing which give algorithmic form to current methods and practices of textual criticism and interpretation. (Buzzetti, S. 62)

Dieser Sichtweise zufolge müsste Markup für Texte unter digitalen Bedingungen dasjenige explizit machen, was in Texten unter analogen Bedingungen immer schon implizit enthalten ist: ihre narrative Strukturierung, die beispielsweise (jedoch nicht ausschließlich) in der grafischen Gestaltung des Textes sichtbar wird. Dass tatsächlich von einer kulturell überlieferten kognitiven Struktur in der grafischen Gestaltung von Texten, einer Struktur, die durchaus schon eine semantische Funktion übernimmt, zu sprechen ist, zeigt sich zum Beispiel an Software, die zumindest die bibliographischen Angaben in Artikeln automatisch von PDF-Format in XML (genauer in TEI) enkodieren kann – und zwar, indem sie grafische Muster der PDF-Vorlage auswertet.[25]

Auch Renears Bewertung der oben genannten unterschiedlichen Textmodelle vollzieht sich auf Grundlage ihrer Funktionalität, das heißt entscheidend ist, welches Modell den einfachsten und funktionalsten Ansatz bietet, um Texte zu produzieren, zu modifizieren und auszutauschen (vgl. Renear 1993). Als den genannten Kodierungsarten eindeutig überlegen wird schließlich digitaler Text herausgestellt, der nach den TEI-Richtlinien in XML (damals noch SGML) augezeichnet ist. Mit TEI ausgezeichnete Texte basieren Renear zufolge auf dem Modell von Text als einer "ordered hierarchy of content objects" (OHCO): Text ist diesem Modell zufolge hierarchisch aufgebaut, die ihn konstituierenden Objekte (zum Beispiel: Kapitel, Absatz) sind ineinander eingebettet. Interessant ist in diesem Kontext, dass Renears Modell tatsächlich aus der Praxis der Textkodierung erwachsen ist und auch damit abzugleichen ist. Dabei zeigt sich, dass die Anwendung der TEI-Richtlinien zu einigen praktischen Problemen führen kann, die auf eben dieses OHCO-Modell zurückzuführen sind, vor allem das "problem of overlapping hierarchies", das sich beispielsweise bei der parallelen Auszeichnung von Abschnitten und der Paginierung nach Vorbild des physischen Mediums (narrative Struktur versus

[25] Vgl. Lopez 2009 und die Betaversion, zu finden unter: http://grobid.no-ip.org.

(Buch-)Repräsentation) oder bei der Enkodierung von Lyrik stellt.[26] "Content objects" wie Sätze und (einem bestimmten Metrum folgende) Verse in Gedichten sind kaum in einer Hierarchie unterzubringen, Enjambements und Verse wiederum nicht. Renear verfeinert daher sein OHCO-Modell: „We have retreated from saying that texts are hierarchical, to saying that perspectives are hierarchical, to saying that perspectives can be decomposed into hierarchical sub-perspectives" (Renear 1993, S. 11). Auf der Konferenz der Association for Computers and the Humanities/Association for Literary and Linguistic Computing (ACH/ALLC) im Jahr 1999 stellte McGann dieses OHCO-Modell (auch in seiner abgeschwächten Form) in Frage:

> McGann instead gave us an example of a reading that was both a performance itself and pointed to the combinatorial possibilities within and around the text. McGann's challenge to Renear was to show how a playful reading of a text was both a new text and that this potential could not be captured easily by an OHCO. (Rockwell 2003)

McGann betont, dass sich insbesondere Lyrik dem OHCO-Modell entzieht (vgl. Renear et al. 1999). Überlappende Strukturen und das Prinzip der Wiederholung (z. B. Reim, Metrik, phonetische Stilfiguren) stellen in der Tat die Essenz von Lyrik dar und zeichnen auch andere literarische Genres aus. Literarische Texte sind nicht nur nach ihren linguistischen Elementen zu beurteilen, sondern nach den Beziehungen, die ihr semiotisches Material bildet. Im „Abstract" zur Konferenz wird McGanns Kritik folgendermaßen wiedergegeben:

> TEI and SGML create, as Renear shows, a certain type of "linguistic" hierarchy, one that privileges text as a container for storing information. But even that linguistic hierarchy is highly specialized (it does not consider, for example, the rhetorical structures that overlap and infect the syntax and semantics). […] Not without reason has the Bergen Wittgenstein project abandoned TEI/SGML as a system for marking up the corpus of Wittgenstein's texts; and the scholars setting out now to "edit" the Peirce archive are well aware that TEI/SGML does not lend itself to an adequate treatment of Peirce's work, and least of all to his existential graphs. "Text" in Kant "is" one thing, but in Peirce it "is" something else again. (Renear et al. 1999)

Die Debatte vermag einige Aspekte hinsichtlich Kodierungsarten, Formaten und Textmodellen aufzuzeigen: Jeder technischen Entscheidung liegt (bewusst oder unbewusst) ein (Text-)Modell zugrunde; in der Praxis kann sich die Konsistenz eines solchen Modells zeigen; häufig werden unterschiedliche Perspektiven und Ebenen –

[26] In den TEI Guidelines P5 gibt es einige Vorschläge zum Umgang mit der „inability to encode overlapping hierarchies" (Burnard/Bauman, S. 634). Dass dies ein schwerwiegendes Problem für die Enkodierung von Texten, zumal von literarischen Texten, darstellt, zeigt sich beispielsweise auch daran, dass es eine eigene Email-Liste zum Thema „overlapping hierarchies and solutions to encode them" und viel Literatur zu diesem Thema gibt (vgl. http://wiki.tei-c.org/index.php/SIG:Overlap). Eine in diesem Kontext interessante Diskussion fand im April 2010 über die von Willard McCarty moderierte Emailliste der „Humanist Discussion Group" statt (Subject: „23.789 inadequacies of markup"; vgl. http://www.digitalhumanities.org/humanist/Archives/Current/Humanist.vol23.txt).

hier die logisch-linguistische mit einer semiotisch-semantischen – miteinander verwechselt, beziehungsweise es entstehen Missverständnisse, weil miteinander nicht kompatible Perspektiven auf ein Phänomen nicht als solche ausgewiesen werden. Dabei bleibt zu bemerken, dass die Auszeichnung literarischer Texte nach TEI im wissenschaftlichen Bereich noch immer alternativlos ist – übrigens führt Renear diese Tatsache stets als ein Argument zur Verifizierung des OHCO-Modells an (vgl. z. B. Renear 2004, S. 225).

Das Problem der "overlapping hierarchies" lässt sich jedoch auch aus einer anderen Perspektive betrachten: im Kontext der Tradition des Kommentars. Markup stellt immer eine Interpretation des Textes dar, wie in den TEI-Richtlinien bemerkt wird:

> Generalizing from that sense, we define *markup*, or (synonymously) *encoding*, as any means of making explicit an interpretation of a text. Of course, all printed texts are implicitly encoded (or marked up) in this sense: punctuation marks, capitalization, disposition of letters around the page, even the spaces between words all might be regarded as a kind of markup, the purpose of which is to help the human reader determine where one word ends and another begins, or how to identify gross structural features such as headings or simple syntactic units such as dependent clauses or sentences. Encoding a text for computer processing is, in principle, like transcribing a manuscript from *scriptio continua*; it is a process of making explicit what is conjectural or implicit, a process of directing the user as to how the content of the text should be (or has been) interpreted. (Burnard/Bauman, S. xxxi)

Dass es angesichts verschiedener Interpretationen "multiple hierarchies" (Witt) gibt, verwundert kaum. Wenn Markup im Sinne des Explizitmachens des im Druckmedium Impliziten, aber auch im Sinne semantischer Auszeichnungen den Text begleitet, entsteht mit verschiedenen Markup-Versionen eines Textes ein Kommentar, eine Rezeptionsgeschichte eines Textes. Diese Enkodierungen, diese Kommentare treten als "a reconfiguration which advances knowledge through a discourse realisable only within an electronic environment" (Schreibman, S. 287) auf und können damit eine lange literarische Tradition auf neue Weise realisieren. Markup spiegelt "the history of the transmission and the interpretation of the text; it reflects its expressive development and conceptual vitality“ (Buzzetti, S. 83). Indem Markup also einem Text Interpretationen hinzufügt, wird ein hermeneutischer Textbegriff wieder hergestellt oder in dieser Form überhaupt erst realisiert.

6. These zur Rekonstruktion des Dokumentbegriffs

Im Zuge einer Rekonstruktion des Dokumentbegriffs unter digitalen Bedingungen kann manches Desiderat eingelöst werden, das gedruckte Texte nicht erfüllen konnten, wobei avantgardistische Schreibweisen in belletristischen Texten diese

digitalen Techniken zum Teil antizipiert haben.

Pédauque sehen Ähnlichkeiten zwischen der Situation, in der es zu Beginn des 20. Jahrhunderts zur „documentarisation“ kam, und der heutigen Situation – mit dem Ziel, die Möglichkeiten einer „redocumentarisation“ aufzuzeigen. Das Ziel der „documentarisation“ war „d'optimiser l'usage du document en permettant un meilleur accès à son contenu et une meilleure mise en contexte. Par extension, on peut dire que documentariser, c'est donner un statut à un texte, le faire 'document'“ (Salaün, S. 6). Ein ähnliches Streben, wenn auch unter veränderten Bedingungen und mit anderen Mitteln, ist im digitalen Zeitalter nach der anfänglichen Emulationsphase zu beobachten: klassische Dokumentationstechniken wie Klassifizieren, Indexieren und Dokumentationssprachen/Thesauri werden von Netzwerkprotokollen wie HTTP, URLs, Web-2.0- und semantisch basierten Technologien abgelöst.[27] Ziel ist es, der Instabilität digitaler Inhalte etwas entgegenzusetzen, die gesellschaftlichen Funktionen, die mit dem Dokumentbegriff verbunden waren, aufrechtzuerhalten, ohne ein bloßes Abbild des analogen Dokuments zu erzeugen, denn die digitale Umgebung bietet neue Möglichkeiten:

> Redocumentariser, c'est documentariser à nouveau un document ou une collection en permettant à un bénéficiaire de réarticuler les contenus sémiotiques selon son interprétation et ses usages à la fois selon la dimension interne (extraction de morceaux musicaux pour les ré-agencer avec d'autres, ou annotation en marge d'un livre suggérant des parcours de lecture différents...) ou externe (organisation d'une collection, d'une archive, d'un catalogue privé croissant les ressources de différents éditeurs selon une nouvelle logique d'association). Dans ce contexte, la numérisation offre des opportunités inédites pour la réappropriation des documents et des dossiers en vue de satisfaire les intérêts de nouveaux bénéficiaires. (Zacklad, S. 282)

Ein Beispiel für die neuen Möglichkeiten zeigen beispielsweise schon heute digitale Editionen. Die Editionswissenschaft war eine Vorreiterdisziplin im Bereich der Digital Humanities. Vor allem die Möglichkeiten zur Visualisierung verschiedener Fassungen, die Möglichkeit, die – häufig durch Herausgeber und Kanonisierungsprozesse verdeckte – Konstruiertheit des literarischen Textes sichtbar zu machen, muss großen digitalen Editionsprojekten Anreiz gewesen sein.[28] Konnten Faksimile-Editionen und lange Anmerkungsapparate zu historisch-kritischen Ausgaben nur ein (zeitliches) Hintereinander und nur unzureichend einen Wechsel

[27] Für eine tabellarischen Vergleich zwischen „documentarisation“ und „redocumentarisation“ vgl. zum Beispiel Salaün 2008, S. 8.

[28] Vgl. z. B. die Stendhal-Ausgabe (http://stendhal.msh-alpes.fr/manuscrits/) oder die Madame-Bovary-Edition (http://bovary.univ-rouen.fr/) für Manuskript-Transkriptions-Zusammenschauen, die Pläne für die digitale Faust-Edition (www.faustedition.net) oder die Zusammenschau von Druckvorlage, Transkription und Enkodierung in XML nach TEI im Deutschen Textarchiv (http://www.deutschestextarchiv.de/).

zwischen vielen verschiedenen Fassungen oder ein Nebeneinander ermöglichen, wird mit synoptischen Editionen im Digitalen nun technisch realisiert, worauf manche Buchedition hinzustreben schien:

> Controlled text annotation, in which any number of layers of information can be merged into one document, allows the several different aspects in a text to be captured and analysed—graphic organization, content, linguistic structure (lexical, morphological, syntactic and semantic) and marks of the stages of diffusion—both vertically and horizontally, in parallel with other text. This is very much what textual studies have been pursuing over the centuries in their task of notating slight differences of terms in the margins of manuscripts, tracing the footprints of previous editors, and drawing intricate interpretation signs for editors to come: the weaving of delicate tangled webs of correspondences between different versions of different documents. (de Sousa, S. 250f.)

Mit den neuen Darstellungsarten sind auch neue Arbeitstechniken und ein neues Selbstverständnis – in diesem Fall von Philologinnen und Philologen – verbunden. Digital Humanities und – spezieller – Computerphilologie sind an deutschen Universitäten zwar inzwischen institutionalisiert,[29] Resonanz besitzt diese Richtung in den Geisteswissenschaften bislang allerdings noch kaum. Die Gründe hierfür werden noch im folgenden Kapitel zu betrachten sein. Abgesehen von digitalen Editionen hat sich für das wissenschaftliche Arbeiten ein ganzer Apparat von Methoden und Werkzeugen ausgebildet, der derzeit beispielsweise im Rahmen von Textgrid Geisteswissenschaftlerinnen und -wissenschaftlern zur Verfügung gestellt werden soll: XML-Editoren, die das Auszeichnen von Texten für spätere – prozessierbare – Analysen beziehungsweise zur Anreicherung von Texten mit weiterführenden Informationen ermöglichen, Lemmatizer zur Optimierung des Recall im Retrieval und vieles mehr.[30]

Pédauque sprechen im Kontext der neuen Möglichkeiten in digitaler Umgebung unter anderem von einer neuen *poeisis* durch die Eingliederung von Repräsentationen einer anderen semiotischen Klasse (also zum Beispiel Abbildungen, Multimedia) in Textdokumente.[31] Für Autorinnen und Autoren literarischer Texte ist diese Art intersemiotischer Poesie schon lange von Interesse. Häufig wird angenommen, dass sich literarische Schreibweisen, dass sich narrative Techniken den Möglichkeiten, die

---

[29] Es gibt inzwischen computerphilologische Lehrstühle an der Universität Hamburg, der Universität Göttingen, der Universität Würzburg und der Universität Wuppertal.

[30] Freie Software zur Textauszeichnung mit zahlreichen Tools bietet z. B. auch die Computerphilologie der Universität Hamburg mit CATMA (http://www.slm.uni-hamburg.de/catma/index.html). TUSTEP, ein frühes und besonders einflussreiches Produkt des Zentrums für Datenverarbeitung der Universität Tübingen, steht inzwischen auch im Rahmen von Textgrid zur Verfügung (http://www.tustep.uni-tuebingen.de/).

[31] „[N]ous utilisons le terme de *poesis* pour désigner la créativité intrinsèque au document, c'est-à-dire la spécificité que le choix des constituants et leur organisation confèrent au document en tant qu'unité cohérente“ (Pédauque 2006, S. 166).

die Technik bietet, anpassen:

> Wenn Bücher demnächst auf allen möglichen Geräten lesbar werden, die gleichzeitig Bilder zeigen, Töne abspielen und Verbindungen zu Internet und anderen Geräten herstellen können, dann wird es nicht mehr lange dauern, bis sich ihrerseits ihre Erzeuger mehr und mehr multimedialer Mittel bedienen, um Werke zu produzieren, die in Gutenbergs Universum keinen Platz mehr finden. Es wird Bestseller geben, die nie als Druckerzeugnis erscheinen, Handyromane in Fortsetzung, die alle lesen, weil alle darüber sprechen (in der U-Bahn von Tokyo eindrucksvoll zu beobachten), undruckbare, multimediale, ständig aktualisierte, reichhaltig animierte Sachbücher, Individualreiseführer oder Enzyklopädien, die kaum noch etwas mit ihren papiernen Vorfahren gemeinsam haben, vernetzte Werke aus Netzwerken von Autoren, verzweigte Geschichten, die vor den Augen des Publikums entstehen, und so vieles mehr, das wir uns heute noch gar nicht vorstellen können. (Neffe 2009)

Obwohl die Beobachtung, dass sich mit der Durchsetzung neuer Technologien neue künstlerische Formen verbreiten, zutrifft, lässt sich die Argumentation auch umkehren. „Die Geschichte jeder Kunstform hat kritische Zeiten, in denen diese Form auf Effekte hindrängt, die sich zwanglos erst bei einem veränderten technischen Standard, d. h. in einer neuen Kunstform ergeben können“, heißt es bei Walter Benjamin (Benjamin, S. 462f.). Intersemiose, die Integration verschiedenartiger semiotischer Systeme (Bild, Ton, bewegtes Bild, Text), drängen seit den Avantgardebewegungen, seit über hundert Jahren, auf veränderte technische Möglichkeiten. Bild und Text sind – wenn auch in unzureichender Qualität und nicht mittels bewegter Bilder – häufig eine semiotische Beziehung in literarischen Texten eingegangen. Neben der Ekphrasis, der reinen Beschreibung von Kunstwerken in literarischen Texten, ist an die Text-Bild-Montagen der Avantgarde-Literatur – insbesondere der Dada-Bewegung – zu denken[32] oder auch in neuerer Zeit an die Texte W.G. Sebalds, in denen Fotografien einen zentralen Bestandteil ausmachen und in einer komplexen Zeichenbeziehung mit sprachlichen Zeichen verwoben werden.[33] Arno Schmidts „Zettels Traum“ demonstriert sehr deutlich, dass Schreibformen an die Grenzen des gedruckten Mediums stoßen und zu neuen technologischen Möglichkeiten hindrängen. Nachdem er 1.350 Großblätter in seiner komplexen Spalten-Technik beschrieben hatte, bemerkt Schmidt:

> Immerhin wird sich das Buch vermutlich nicht mehr setzen lassen, sondern – das MaschinenManuskript wird fotomechanisch vervielfältigt werden müssen. Eine Not aus der ich – ne Tugend insofern gemacht habe als ich auch – mir Zeichnungen am Rande erlaubt habe, ohne Bildvorlage die mich anregten. (Schmidt, S. 4)

Natürlich war, wie die Herausgeber der Bargfelder Ausgabe 1992 „Zum Satz der

[32] „Ein Blick auf die typographischen Standards in den Printmedien und der Werbung nach der Jahrhundertwende zeigt, daß die revolutionäre Leistung Dadas vor allem darin besteht, Textformen aus den Massenmedien ins Kunstsystem 'importiert' zu haben“ (Wehde, S. 400). Bei dem importierten Material handelt es sich nicht nur um Textformen, sondern auch um Bilder.

[33] Vgl. z. B. W. G. Sebalds „Die Ausgewanderten. Vier lange Erzählungen“ (1992), „Die Ringe des Saturn“ (1995) oder „Austerlitz“ (2001).

späten Typoskripte Arno Schmidts“ anmerken, „das schließlich praktizierte photographische Faksimileverfahren eine aus technischen und finanziellen Schwierigkeiten geborene Notlösung“ – das „Einmontieren von Zeichnungen und Bildern in gesetzten Text ist mit den heutigen technischen Möglichkeiten keine Schwierigkeit“ (aus der Vorbemerkung zur Ausgabe „Zum Satz der späten Typoskripte Arno Schmidts“, Arno Schmidt Stiftung).

Avantgardistische Formen lassen Neuartiges in den Bereich des Denkmöglichen vordringen und antizipieren derart auch technische Möglichkeiten. Umso bedauerlicher, wenn diese Möglichkeiten nach der Realisierung ihrer technischen Bedingungen nicht genutzt werden, sondern die alten Restriktionen in die digitale Welt übersetzt werden:

> For over two computer decades we have imitated paper on screens (extremes being Adobe Acrobat and the HTML page). What we really need is the opposite -- to represent digitally the literary forms of connection which could not be represented before. (Nelson 1999)

Es ist nicht verwunderlich, dass Nelson diese Verwirklichung noch immer in seiner Software Xanadu sieht, an der seit den 60er Jahren des vorigen Jahrhunderts zahlreiche Menschen arbeiten und von der auch heute nur eine äußerst minimalistische Demo-Version herunterzuladen ist.[34] Dabei sind Nelsons Kritik und auch seine Ideen, wie sich Dokumente aufeinander beziehen können müssten, reizvoll – wenn auch äußerst komplex in ihrer Realisierung: Dokumente sollen in „transpointing windows“ als parallele Visualisierung zum Vergleich nebeneinander gestellt werden können, wobei auf Mikrostrukturen in Dokumenten referenziert werden kann; dabei wird zwischen „links“ und „transclusions“ unterschieden, wobei Letztere zur Nachnutzung von Teilen anderer Dokumente – ähnlich dem früheren Zitieren – gedacht sind, allerdings verbunden mit der Visualisierung der Nachnutzung, also dem Verweis auf den Originalkontext; es existieren auch reziproke Links; ein eigenes „transcopyright“ mit Micropayment-System ermöglicht eine einfache relativ offene Nutzung von Dokumenten; und schließlich gibt es ein Versionierungsmanagement (vgl. Nelson 2007).

Xanadu beinhaltet im Sinne der Nachnutzung von Dokumenten auch Ansätze für kollaboratives Arbeiten, dessen Förderung im geisteswissenschaftlichen Bereich auch eines der Ziele von Textgrid ist. Im literarischen Kontext hat „Sympoesie“ von der Frühromantik über die experimentelle Dichtung des 20. Jahrhundert bis heute

[34] Download unter http://www.xanadu.com/ Zur Geschichte des Xanadu-Projekts vgl. Wolf.

immer eine Rolle gespielt. Das WWW schwächt den Autorbegriff: der individuelle Autor verliert zumindest im Kontext von Webseiten, weniger in wissenschaftlichen Emulationsszenarien, an Gewicht: das WWW ist ein Netz von Diskursen. Autorenkollektive wie Roger T. Pédauque stellen den Autorbegriff, der in den letzten zwei Jahrhunderten zu einem zentralen kulturellen Konzept geworden ist – wie beispielsweise auch die eingangs referierte Debatte um Hegemanns Roman zeigt –, auf die Probe und schließen damit an Experimente aus dem literarischen Bereich an, die nicht alle eine so lange Lebzeit wie Oulipo, „L' **Ou**vroir de **Li**ttérature **Po**tentielle“, hatten.

Neben kollaborativem Schreiben lässt die digitale Umgebung insbesondere in Bezug auf Intertextualität ein ganz neues Szenario entstehen. Stehen Bücher als physisch isolierte Entitäten im Regal, entwickelt digitaler Text Synapsen (vgl. Cherval). Für Kevin Kelly sind Links und Tags sogar „the most important inventions of the last 50 years“:

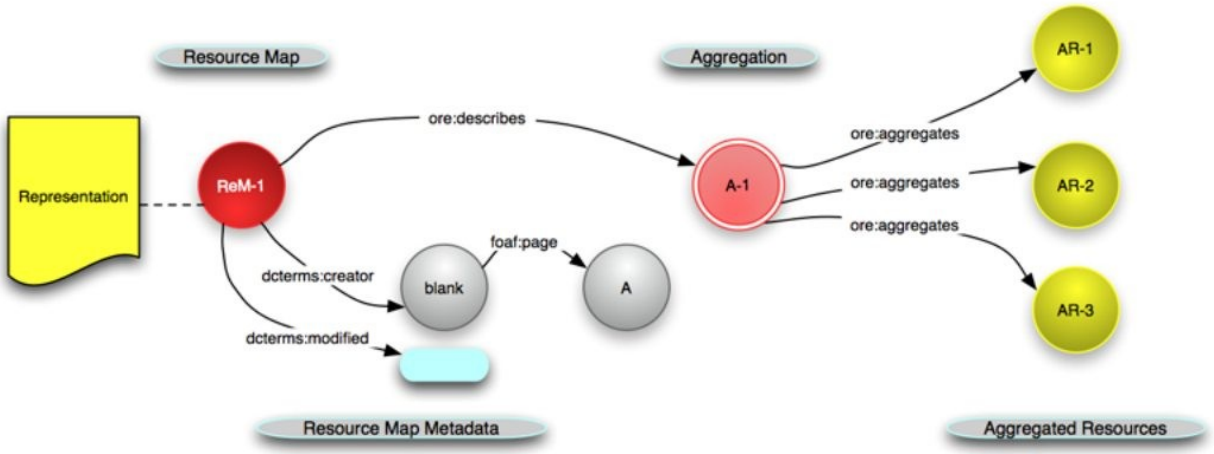
> Turning inked letters into electronic dots that can be read on a screen is simply the first essential step in creating this new library. The real magic will come in the second act, as each word in each book is cross-linked, clustered, cited, extracted, indexed, analyzed, annotated, remixed, reassembled and woven deeper into the culture than ever before. In the new world of books, every bit informs another; every page reads all the other pages. (Kelly)

Die Frage, die sich in dieser Situation stellt, zielt zugleich ins Herz von Pédauques „redocumentarisation“-Bemühungen: Wie ist ein stabiles und trotzdem für die neuen Techniken flexibles Dokument unter digitalen Bedingungen herzustellen? Eine mögliche Richtung stellen dabei die bereits genannten semantisch basierten Technologien dar. Mittels RDF lassen sich Bezüge zwischen Informationsobjekten bilden. Wenn es möglich wäre, bestimmte RDF-Strukturen, zum Beispiel in Form einer Aggregation, zu stabilisieren, könnte diese Aggregation dokumentähnliche Funktionen mit neuen Funktionalitäten übernehmen.

Als eine Reaktion auf diese Entwicklungen kann die Open Archives Initiative Object Reuse and Exchange (OAI-ORE) verstanden werden, die auf der Basis von RDF Standards für die Beschreibung und den Austausch von Ressource-Aggregationen definiert. Erwachsen ist OAI-ORE eben aus dem Problembewusstsein, dass es keinen Standard zur Identifizierung der derzeit im WWW sehr ephemer vorliegenden Aggregationen gibt: „In essence, the problem is that there is no standard way to describe the constituents or boundary of an aggregation, and this is what OAI-ORE aims to provide“ (Lagoze/Van de Sompel). Indem eine URI für die Aggregation

vergeben wird und es zusätzlich eine „Resource Map“ gibt, die eine eigene URI hat und die – für Maschinen und Menschen lesbar – zur Beschreibung der Aggregation dient (etwa welche Ressourcen Teil der Aggregation sind und welche ihr nur verbunden sind), wird – um die topografische Metaphorik zu übernehmen – eine Grenzziehung vorgenommen und eine Karte des Geländes erstellt.

Figure 3: The Aggregation A-1 aggregates three Resources and is described by Resource Map ReM-1



(Abbildung 3 aus Lagoze/Van de Sompel)

Aggregationen könnten beispielsweise verschiedene Versionen eines Textes bilden. Im Bereich der Naturwissenschaften verdrängen Publikationen von Primärdaten zusammen mit Prozessierungsmethoden bereits konventionell in Textform verfasste Publikationen. Insbesondere für empirische Wissenschaften, wie beispielsweise Soziologie, ist jedoch auch die Publikation von Primärdaten zusammen mit dem Artikel, der die Ergebnisse auswertet, von Interesse – als eine Aggregation, ein Dokument. Für die Geisteswissenschaften stellen diese Entwicklungen jedoch einen sehr viel fundamentaleren Umbruch dar: Die Bedingungen der Produktion und Distribution verändern sich ebenso wie die grundlegenden Bedeutungsmodi von Dokumenten als komplexen Zeichen (vgl. Gradmann 2010).

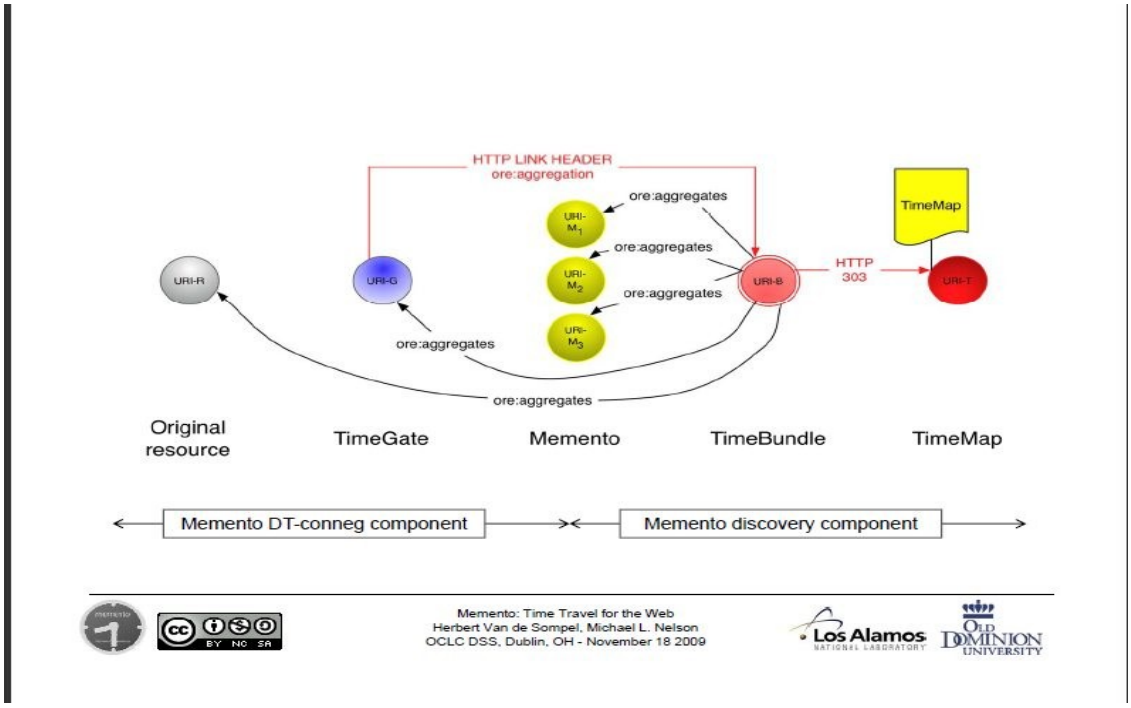
Als digitale Primärdaten im Bereich der Philologien können die Primärtexte gelten, ohne die ein Verständnis der Sekundärliteratur zumeist überhaupt nicht möglich ist. Als angereicherte Primärtexte, die bereits in den Bereich der Sekundärliteratur übergehen, könnten semantisch getaggte Texte verstanden werden. Digitale Texte sind eben aufgrund der unterschiedlichen Interpretationsschichten, die sie tragen können, potenziell reicher als gedruckte Texte, sie erfüllen viel eher das, was Genette

in seinen „Palimpsesten“ als Metatext-Dimension bezeichnet: die Schichten von Rezeption, Interpretation, Fortschreibungen (Genette). Als Aggregation verschiedenartig getaggter Texte erhielte man so eine ganze Rezeptionsgeschichte des Primärtextes. Im nicht-wissenschaftlichen Bereich ließen sich Aggregationen für jegliche Art von Sammlungen im Sinne von Anthologien realisieren, die sehr viel schneller und (zielgruppen-)genauer realisierbar wären, beziehungsweise den Rezipientinnen und Rezipienten die Möglichkeiten bieten würden, selbst eigene Zusammenstellungen zu treffen (und zu teilen).

Auch die Ansätze des Semantic Publishing, das derzeit vor allem im Bereich des wissenschaftlichen Publizierens erprobt wird, versuchen die Technologien des Linked-Data-Kontinuums in Zukunft besser zu nutzen. Schon 2001 erwarteten Berners-Lee/Hendler eine baldige grundlegende Veränderung des wissenschaftlichen Publizierens durch semantisch basierte Technologien: „In the next few years, we expect that tools for publishing papers on the web will automatically help users to include more of this machine-readable mark-up in the papers they produce“ (Berners-Lee/Hendler). Dementsprechend versteht Shotton unter Semantic Publishing „anything that enhances the meaning of a published journal article, facilitates its automated discovery, enables its linking to semantically related articles, provides access to data within the article in actionable form, or facilitates intergration between papers“ (Shotton 2008, S. 2). Beispiele für solche Funktionalitäten sind Hyperlinks (zum Beispiel zu den Institutionen der Wissenschaftlerinnen und Wissenschaftler, die den Artikel geschrieben haben), semantisches Markup mit Links zu weiterführenden Informationen, interaktive Abbildungen, eine Literaturliste, die exportiert und weiter bearbeitet werden kann, Dokumentanalysen in Form von Tag-Bäumen und Tag-Clouds, Dokumentstatistiken inklusive numerischer Analysen der Zitationen. Ein Problem von semantischem Markup ist, dass selbst ein auf eine bestimmte Disziplin begrenztes Schema, ein kontrolliertes Vokabular, kompliziert zu erstellen und vor allem nicht unveränderbar haltbar ist. Semantisches Markup im großen Stil – im Sinne eines Semantic Web – erscheint in diesem Kontext als geradezu naiv.

Fraglich bleibt bei all diesen Bemühungen, wie der Inhalt der jeweiligen Ressource stabilisiert werden kann, ohne PDF-ähnliche Techniken zu nutzen, die den dynamischen Charakter des WWW konterkarieren. Ein Ansatz ist, Zeitstempel für

Datenaggregationen zu vergeben, das heißt, Webressourcen und ihre unterschiedlichen Repräsentationen in verschiedenen Versionen entlang des Zeitvektors zurückzuverfolgen. Diesen Ansatz verfolgen Van de Sompel und (Michael L.) Nelson mit ihrem Memento-Projekt. Bislang besitzt HTTP beziehungsweise das WWW keine zeitliche Dimension.[35] Zeitlich zurückliegende Versionen von Repräsentationen bestimmter Ressourcen können zwar mit einigem Aufwand in verschiedensten Web-Archiven gefunden werden, die Dokumente werden jedoch meist wie in zeitlichen Inseln angezeigt, ihre Links führen nicht zu dem damaligen Zustand der vernetzten Ressourcen, sondern zum aktuellen. Das Projekt Memento simuliert eine Zeitreise: das WWW soll – mittels Zeitvektor in der Content Negotiation zwischen Client und Server und durch eine API für Archive – in verschieden lange zurückliegenden vergangenen Zuständen zugänglich werden. Die verschiedenen Versionen könnten durch ORE zu einer Aggregation verschmolzen werden (vgl. Van de Sompel 2009).



(Die Abbildung entspricht der Folie 63 der „Slides of Herbert Van de Sompel's Memento presentation“ (Van de Sompel).)

Mittels Einführung eines Zeitvektors wäre das Referenzieren auf eine eindeutig zu bestimmende Version möglich – als Entsprechung einer traditionellen

[35] Dass dem WWW ein Zeitvektor fehlt, hat dem W3C zufolge pragmatische Gründe: „Requiring a URI owner to publish a new URI for each change in resource state would lead to a significant number of broken references. For robustness, Web architecture promotes independence between an identifier and the state of the identified resource“ (Jacobs/Walsh 2004).

Dokumentfunktionalität. Damit sind zwar noch nicht Fragen nach Langzeitarchivierbarkeit gelöst, aber zumindest der Weg zur Langzeitverfügbarhaltung geebnet.

3 Digitale Belletristik als Primärliteratur für philologische Forschung in Wissenschaftlichen Bibliotheken

Die digitalen Bedingungen besitzen das Potenzial, die wissenschaftlichen Arbeitsmethoden, Publikationsweisen und -wege fundamental zu verändern. Hat in den Naturwissenschaften längst mit Virtuellen Forschungsumgebungen, elektronischem Publizieren in kommerziellen oder Open-Access-E-Journals, vermehrt auch von Primärdaten, und dem Archivieren in Repositorien ein Übergang des wissenschaftlichen Arbeitens und Publizierens vom Analogen ins Digitale stattgefunden, ist eine derartige Entwicklung in den Geisteswissenschaften nur ansatzweise erkennbar. Ein Grund für diese Differenz ist in den unterschiedlichen Wissenschaftskulturen und ihrem jeweiligen Verhältnis zur Sprache zu suchen. Während in den empirischen Wissenschaften nicht-diskursive Ergebnisse in wissenschaftliche Kommunikationsprozesse transportiert werden, sind in den hermeneutischen Wissenschaften der Erkenntnis- und Kommunikationsprozess im semiotischen System der Sprache untrennbar miteinander verbunden, insbesondere in den Philologien (vgl. Gradmann 2004). Für die Naturwissenschaften ergeben sich im Vergleich zum gedruckten Medium im Digitalen offensichtlich neue Möglichkeiten der Publikation von nicht-sprachlichem Material; für textbasierte Publikationen werden größtenteils kürzere Formen wie Zeitschriftenartikel gewählt. In den Geisteswissenschaften gestaltet sich die Situation anders. Hier steht die gedruckte Monographie und somit ein traditioneller Dokumentbegriff im Zentrum. Schon alleine aufgrund ihres Umfangs und ihrer Komplexität wird die geisteswissenschaftliche Monographie kaum als elektronische Publikation rezipiert. Wenn das Dokument als monolithisches komplexes Zeichenobjekt grundlegend zur Kultur der textbasierten Wissenschaften gehört, verwundert es nicht, dass sich in diesem Bereich unter digitalen Bedingungen Emulationsszenarien beständiger zeigen als in den empirischen Wissenschaften:

> [M]ost humanities researchers still use digital resources as if they were print. Once they have their sources neatly printed out and sorted in piles, they conduct the real research themselves. And the ways in which we have constructed our thematic research collections lend themselves to that treatment. We still present text in codex-like ways to be read within a narrative context. What we have created is an easily-accessible way to find and print out texts (Schreibman 2009, Abschnitt 35).

Dennoch bieten sich gerade für textbasierte Wissenschaften im Digitalen, verstanden

als eine „datenzentrierte Wissenschaft“, die „Experiment, Theorie und Simulation zusammenführt“ (Neuroth et al., S. 162), neue Möglichkeiten, insbesondere im Umgang mit den Quellen, das heißt im Fall der Philologien: vor allem mit den Primärtexten. Denn gerade für die Linguistik und die Philologien eröffnet das Charakteristikum der Diskretheit des digitalen Dokuments neuartige Verfahrensweisen der Textanalyse (vgl. These 3) – vor allem in Hinblick auf große Textmassen, digitale Korpora. Durch linguistische Erschließung – beispielsweise lemmatisierte Wortlisten, in denen Worte auf ihre Grundformen zurückgeführt werden, was insbesondere aufgrund des Flexionsreichtums der deutschen Sprache von fundamentaler Bedeutung ist –, aber auch durch Verfahren des Information Retrieval, die Suchergebnisse gewichten und ordnen, hat sich die philologische Forschung in den letzten Jahren gewandelt. Befördert wird dieser Prozess zudem durch quantitative Methoden der Textuntersuchung, die stylometrische und strukturalistische Analysen – wie beispielsweise die Anwendung computerphilologischer Verfahren auf narratologische Fragestellungen – vereinfachen und in Bezug auf ein großes Textkorpus überhaupt erst ermöglichen, durch differenzierte Konkordanzprogramme, die einen Vergleich von (Einzel-)Texten erlauben, und schließlich durch neue Visualisierungstechniken im Kontext digitaler Editionen. Allerdings finden computergestützte Arbeitsmethoden bislang im geisteswissenschaftlichen Bereich vornehmlich in der (Computer-)Linguistik, der Editionsphilologie und der Mediävistik Anwendung, in der – hermeneutisch orientierten – Literaturwissenschaft spielen sie bislang kaum eine Rolle (vgl. z. B. Neuroth et al., S. 164), möglicherweise weil mit ihnen „die Aufteilung der wissenschaftlichen Welt in die 'zwei Kulturen', hier die auf Verstehen ausgerichteten Geisteswissenschaften, dort die nomothetischen Naturwissenschaften' (Lauer 2009) verschwimmt. Auch Konkordanzuntersuchungen mit Hilfe von Computern werden in den Philologien, die sich mit neueren Texten beschäftigen, bislang nicht in größerem Maßstab angewendet. Die genannten Verfahren gehen von einem Begriff von Text und Dokument aus, der nicht mehr mit dem des Printzeitalters in Übereinstimmung zu bringen ist. Statt des abgegrenzten Primärtextes, der in Buchform vorliegt, operieren Textmining-Verfahren zunächst mit Daten, mit Ketten alphanumerischer Zeichen, die von ihrer Repräsentation unabhängig prozessiert werden – wobei dieser Textbegriff für viele dieser Verfahren nicht ausreichend ist (vgl. These 5). Nichtsdestotrotz ist die Durchdringung der Dokumentgrenzen im Kontext der

Schaffung und Prozessierung großer Textkorpora eine Bedingung für die Arbeit mit digitalem Text. Anders als bei Text in gedruckter Form, für den der Dokumentbegriff mit all seinen kulturell verbürgten Funktionalitäten greift, rückt im Bereich der Digital Humanities oder der Computerphilologie der Text unabhängig vom Medium (denn potenziell kann alles Analoge auch ins Digitale überführt werden) in den Fokus (vgl. These 3 und 4). Meister definiert Computerphilologie, indem er sie als wissenschaftliche Praxis des Forschungsfelds Humanities Computing versteht, die sich „mit Aspekten der rechnergestützten Bearbeitung traditioneller philologischer Gegenstände“ befasst (Meister 2003, S. 6). Im Vergleich zur traditionellen Philologie sei der Gegenstandsbereich ausgeweitet:

> Diese Gegenstände sind in unserem Fall nun nicht nur Bücher, sondern alle Formen, in denen verschriftlichte Sprache erscheinen kann, mithin *Texte*. Der Textbegriff ist dabei nicht essentiell oder material bestimmt: es ist egal, ob ein Text als Buch, als Zeitungsartikel, als Website u.ä.m. erscheint, solange er nur eine symbolische Repräsentation von sprachlichen Äußerungen liefert. Mit anderen Worten, Texte sind definiert über ihre semiotische Funktion als sprachliche Bedeutungsträger. *Gegenstand der Philologie* aber sind damit in allererster Linie die *Fragen, die sich auf Texte* beziehen: Fragen also, die die Genese, Authentizität, Bedingungshaftigkeit, Regularität und Funktionalität von Texten als Bedeutungsträgern betreffen. (Meister 2003, S. 7)

Wenn der Textbegriff im semiotischen Sinne gegenüber dem materiell fundierten Text- und somit dem Dokumentbegriff für die Computerphilologie an Bedeutung gewinnt, scheint damit jedoch zugleich auch die Differenz zum informatischen Verständnis von Text als Datenform auf. Der Textbegriff der Computerphilologie impliziert eine Potenzierung der bereits in der traditionellen Philologie vorliegenden Repräsentationsbedingungen, so dass „die rechnergestützte Bearbeitung traditioneller philologischer Gegenstände“ als „Methode zur Untersuchung der Repräsentation einer Repräsentation“, in der Text „zu einem in einer neuen, formaleren Symbolsprache repräsentierten System wird“, verstanden werden muss (Meister 2003, S. 8). Um Text prozessierbar zu machen, muss mit der Umwandlung in alphanumerischen Zeichensatz und letztlich in Binärkodierung eine weitere Signifikantenebene, ein weiteres Symbolsystem eingesetzt werden.

Die für die Verfahren essentielle Strukturiertheit von Daten, die den Binärkode letztlich erst in Einklang zum Textbegriff bringen kann, erfolgt durch computer- und menschenlesbare Auszeichnungssprachen – eine Begrifflichkeit mit einer aufschlussreichen ökonomischen Metaphorik:

> Auszeichnung heißt in diesem Zusammenhang: die explizite und eindeutige Beschreibung von relevanten formalen und inhaltlichen Textmerkmalen. Erst diese Explikation nämlich macht aus Texten genuine *Textdaten*. Die Methode der elektronischen Textauszeichnung nennt man *tagging* (ein übertragener Gebrauch des Englischen *tag*, was soviel wie *Preisschild* oder

> *Anhänger* heißt). Auch das Produkt des Prozesses wird mit einem Wort bezeichnet, das man eigentlich aus dem Handel kennt: *Mark-up*. Der Begriff bedeutet im Literalsinn soviel wie Aufschlag, Mehrwert oder Gewinnspanne. Die ökonomische Metapher ist durchaus passend, denn auch für den Geisteswissenschaftler erhält ein Text, der sich durch *tagging* in *Text+Mark-up = Textdaten* verwandelt, einen ganz entscheidenden Mehrwert. Aber wie in der ökonomischen so schlägt sich auch in der wissenschaftlichen Praxis Mehrwert nur dann in Tauschwert nieder, wenn man sich auf eine gemeinsame Währung einigen kann.“ (Meister 2003, S. 16)

Im Zuge der Bemühungen um eine Standardisierung der Textauszeichnung ist die Text Encoding Initiative (TEI) von großer Bedeutung. Bereits 1987 traf sich eine Gruppe von Philologen, die „Guidelines“ für die Kodierung von (literarischen) Texten erarbeitete, wobei auf den 1986 verabschiedeten ISO-Standard zur Definition von Markup-Sprachen SGML (Standard Generalized Markup Language) zurückgegriffen wurde. Die erste Version der Guidelines, TEI P1, wurde 1990 publiziert (vgl. Burnard/Bauman zum „Historical Background“, S. xxviii ff.). Mit der Entwicklung und Verbreitung von XML wurde TEI auf XML, eine vereinfachte Teilmenge von SGML, umgestellt. TEI stellt den textbasierten Wissenschaften mehr als 500 Elemente zur Auszeichnung von Texten zur Verfügung und hat sich innerhalb der Geisteswissenschaften zu einem De-facto-Standard entwickelt, den fast alle digitalen Textarchive übernommen haben.[36] Die Text Encoding Initiative ist auch deswegen eine Besonderheit, weil sie eine „large-scale contribution from humanities computing to a shared technical agenda of widespread interest and influence“ (McCarty, S. 205) darstellt. Sie entstammt den Geisteswissenschaften selbst; ihre Richtlinien sind das Ergebnis eines langen wissenschaftlichen Diskussionsprozesses und sind als ein Versuch zu verstehen, „das zentrale Wissen über die literaturwissenschaftliche und linguistische Textauszeichnung zu sammeln“ (Jannidis 2007, S. 29). Sie stellen eine Kondensation von philologischem Wissen dar, das daher auch niemals endgültig fixiert ist, sondern stets den Debatten der Zeit gemäß angepasst werden muss – oder wie Schreibman es ausdrückt: „TEI was not simply a format for exchanging documents, but a way of representing, describing, and understanding textual evidence“ (Schreibman 2009, Abschnitt 1). Diese Anpassung wird zwar in den verschiedenen Versionen der Guidelines vollzogen, allerdings stellen die TEI

[36] Z. B. das Women Writers Projekt, das bereits seit den 1980er Jahren besteht (http://www.wwp.brown.edu/), das British National Corpus Project, das seit Anfang der 1990er Jahre elektronische Texte TEI-konform kodiert (http://www.natcorp.ox.ac.uk/), im deutschsprachigen Raum das Deutsche Textarchiv (http://www.deutschestextarchiv.de/), das seit 2007 mit Förderung der DFG an der Berlin-Brandenburgischen Akademie der Wissenschaften angesiedelt ist, und Das digitale Wörterbuch der deutschen Sprache des 20. Jahrhunderts (DWDS,http://www.dwds.de/), dessen Kernkorpus auch belletristische TEI-getaggte Texte enthält. Textgrid hat mit zeno.org eine große digitale Sammlung deutschsprachiger Texte erworben, für die eine Konvertierung in TEI und eine tiefere Auszeichnung geplant ist (vgl. http://www.textgrid.de/abstimmung.html).

Guidelines schon allein aufgrund ihres Umfangs kein besonders flexibles System für Bedeutungswandel dar.

Ein zentrales Ziel der „Guidelines for Electronic Text Encoding and Interchange“ ist es, die Interoperabilität von digitalen Texten zu gewährleisten, ohne sich von bestimmten Programmen und Betriebssystemen abhängig zu machen – und somit auch die Voraussetzung zur Langzeitarchivierung digitaler Texte zu schaffen. Mit TEI ausgezeichneter Text kann mit jedem Programm, das XML lesen kann, dargestellt und verarbeitet werden. Interoperabilität ist auch die Voraussetzung dafür, verschiedene Editionen, die TEI verwenden, zu Recherchezwecken zu einem großen Korpus zusammenzustellen. Als weiteres Ziel der TEI kann die Entwicklung von Tools für TEI-getaggte Texte gelten, die jedoch etwas schleppend verläuft (vgl. Rockwell; Schreibman 2009).[37] Mit TEI sollte jedoch vor allem eine Art best practice der Textauszeichnung geschaffen werden. Dass die aktuelle Version der TEI-Richtlinien (TEI P5) mit ihren 23 Kapiteln in der PDF-Version 1420 Seiten umfasst und mehr als 500 Elemente anbietet, deutet bereits auf ein Problem der TEI hin: die Dokumentation des philologischen Weltwissens wird – allein aufgrund der Basierung auf XML – in ein hierarchisches System eingepasst, das schon durch seinen bloßen Umfang und seine Detailliertheit viele Anwender überfordert.

> Similar resistance from scholars to the textual ontologies provided by the Text Encoding Initiative suggests that resistance is neither accidental nor perverse, nor can it be a property of the medium. Clearly it points to the felt need for maximum flexibility in the design of new scholarly forms. [...] such ontologies are 'as if' constructs whose scholarly purpose is to raise questions, to say what might be and not what is. Practical considerations, however point the need for standard, higher-level categories, not only to serve the impatient or other-minded scholar but also to open up the resource to wider audiences (McCarty, S. 124f.)

Romary zufolge können die TEI-Guidelines nicht nur das Gefühl „*It is too big“*, sondern gleichzeitig auch den Eindruck „*It is not enough*“ auslösen, nicht hinreichend passgenau für das spezifische Projekt zu sein (Romary 2009, S. 8). Obwohl Romary, seit 2010 TEI Council Chair, diese „Vorurteile“ zu entkräften sucht, sind sie doch sprechend für die grundlegende Herangehensweise der TEI, die eher dem bibliothekarischen als dem WWW-Denken entsprungen zu sein scheint, das heißt eher einer hierarchischen präskriptiven Wissensorganisation entspricht als der rhizomartigen experimentellen und postskriptiven Struktur, die sich beispielsweise in RDF-Vernetzungen zeigt. Dass mit TEI Lite eine vereinfachte, reduzierte Fassung

[37] Roma enthält ein Set von Tools (http://wiki.tei-c.org/index.php/Roma). Einzelne Funktionalitäten bieten z. B. die Versioning Machine (www.v-machine.org) oder The UVic Image Markup Tool Project (http://tapor.uvic.ca/~mholmes/image_markup/).

von TEI zur Verfügung steht (Burnard/Sperberg-McQueen 2006), die für Bibliotheken [38], wissenschaftliche Verlage[39] beziehungsweise für Anbieter, die für Verlage digitale Textfassungen herstellen, angepasst wird, deutet an, dass die TEI-Richtlinien für viele Beteiligte schlicht ein zu großes und dementsprechend unflexibles Regelwerk darstellen.

Um die Veränderungen, die ein mit TEI getaggter Text für den Text- und Dokumentbegriff nach sich zieht, betrachten zu können, ist es aufschlussreich, die Strukturierung der digitalen Texte nach den TEI-Richtlinien zu betrachten. Ein grundsätzliches Gestaltungsmerkmal von TEI-Texten besteht in der Unterscheidung zwischen dem TEI-Header und dem TEI-Textkörper. Der TEI-Header beschreibt den enkodierten Text. In der aktuellen Version P5 der TEI-Richtlinien heißt es, dass der TEI-Header „an electronic analogue to the title page attached to a printed work“ sei und eine derartige Dokumentation sowohl für Wissenschaftler, die die Texte nutzen, als auch für Software, die die Texte prozessiert, und schließlich für die Katalogisierung in Bibliotheken und Archive notwendig sei (Burnard/Bauman, S. 17). Die Analogie zum gedruckten Werk erscheint in diesem Fall jedoch kaum passend, eben weil der Header auch maschinenlesbar ist und wichtige Informationen für die Prozessierung der Texte liefert. Der TEI-Header enthält außerdem – je nach Projekt – häufig bereits eine sehr ausführliche Dokumentation des enkodierten Textes beziehungsweise seiner gedruckten Vorlage und des Projekts, in dem er entstanden ist. Ein idealer TEI-Header orientiert sich an folgender Struktur:

```
<teiHeader>
<fileDesc>
<!-- ... -->
</fileDesc>
<encodingDesc>
<!-- ... -->
</encodingDesc>
<profileDesc>
<!-- ... -->
</profileDesc>
<revisionDesc>
<!-- ... -->
</revisionDesc>
</teiHeader>
```

[38] Vgl. die TEI Special Interest Group on Libraries (http://www.tei-c.org/Activities/SIG/Libraries/index.xml) und ihre „Guidelines for Best Encoding Practices“ (http://www.diglib.org/standards/tei.htm).

[39] Vgl. die TEI Special Interest Group on Scholarly Publishing (http://www.tei-c.org/Activities/SIG/Publishing/index.xml) und ihre „Recommended Practices“ (http://wiki.tei-c.org/index.php/Recommended_Practices).

(Burnard/Bauman, S. 18)

Die „file description“ (<fileDesc>) enthält die bibliographische Beschreibung der Datei und ist verpflichtend: „The file description element of the TEI Header has [...] been closely modelled on existing standards in library cataloguing“ (Burnard/Bauman, S. 21). Die anderen Komponenten sind fakultativ: die „encoding description“ enthält Angaben zur Beziehung zwischen elektronischem Text und seiner (analogen) Quelle; die „profile description“ Angaben zur Sprache des enkodierten Textes und dessen Entstehungskontext oder beispielsweise auch Schlagworte; die „revision history“ dokumentiert die unterschiedlichen Versionen der Datei. Auf diese Weise kann eine ausführliche Dokumentation der Datei entstehen, die weit über die bibliographischen Angaben eines Titelblatts hinausgeht. Die Unterscheidung zwischen Header und Textkörper sowie die Tatsache, dass jeder TEI-Text stets seine Metadaten mit sich führt, macht ihn zu einem „completely autonomous digital object that can be archived, transferred or manipulated independently of any extra third party information“ (Romary 2009, S. 12f.).

Der enkodierte Text selbst befindet sich im <text>-Element. Vor der Enkodierung muss eine Entscheidung über den Granularitätsgrad des TEI-getaggten Textes getroffen werden, die gemäß der anschließenden Nutzung des digitalen Textes gewählt werden sollte:

> It is actually important to consider, prior to the starting of any encoding project – even more than the choice of encoding scheme – to which purpose the encoding activity is intended. The trade-off here is to provide a balance between the effort put into accurately encoding many aspects of the source text (or the natively created digital text when applicable) and the actual benefit that can be derived from this encoding from the point of view of legibility or processing. One of the main risks here is to tend towards an encoding over-kill, whereby a lot of effort is put into encoding detailed phenomena in a text, which no user other than the encoder himself will ever need to use. (Romary 2009, S. 2)

Grundsätzlich werden – beispielsweise in „TEI Text Encoding in Libraries“ – fünf Granularitätslevel von „Fully Automated Conversion and Encoding“ und „Minimal Encoding“ über „Simple Analysis“ bis zu „Basic Content Analysis“ und „Scholarly Encoding Projects“ unterschieden.[40] Eine minimale Auszeichnung geht auf die Strukturierung des Textes durch Absätze und Abschnitte ein (mittels <p>- und -Elementen) sowie auf Titel und Zwischenüberschriften (<head>). Einige Elemente sind genreabhängig: beispielsweise die Auszeichnung von Versen beziehungsweise Versgruppen (<lg>) oder die Einführung eines Sprechers in einem Drama (<speaker>).

[40] Vgl. http://www.diglib.org/standards/tei.htm#levels

Dass TEI als Auszeichnungssprache für Digitalisate entwickelt wurde und bis heute zum größten Teil auch zur Auszeichnung von belletristischen Inhalten vor allem von Buchdigitalisaten genutzt wird, zeigt sich unter anderem an der Struktur des <text>-Elements in der XML-Datei: sehr grob besteht sie aus <front>, <body>, <back>.

Das <front>-Element sollte die einleitenden Teile eines Textes enthalten. Obwohl in den „Guidelines“ angegeben wird, dass mit „*front matter*“ die „distinct sections of a text (usually, but not necessarily, a printed one), prefixed to it by way of introduction or identification as a part of its production“ (Burnard/Baumeister, S. 166) gemeint sind, ist diese Art der Strukturierung eines Textes stark an die Tradition des Kodex gebunden. Das Deutsche Textarchiv, dessen Ziel es ist, „einen disziplinübergreifenden Kernbestand deutschsprachiger Texte aus der Zeit von ca. 1650 bis 1900 nach den Erstausgaben zu digitalisieren und als linguistisch annotiertes Volltextcorpus im Internet bereitzustellen“,[41] und das dementsprechend viel Wert auf die Repräsentation der äußeren Gestaltung des Originals legt, bietet neben im Browser lesbaren Text und Faksimile-Bildern auch den Text in XML an. So wird beispielsweise bei der Kodierung eines Digitalisats von Theodor Fontanes „Wanderungen durch die Mark Brandenburg“ in der XML-Datei auch die Repräsentation des Einbands mit in das <front>-Element geholt:

```
<front>
<pb facs="#f0001" />
<pb facs="#f0002" />
<pb facs="#f0003" />
 <pb facs="#f0004" />
<pb facs="#f0005" />
<pb facs="#f0006" />
 <pb n="[I]" facs="#f0007" />
<titlePage type="halftitle"><docTitle><titlePart type="main"> <hi
rendition="#b">Wanderungen</hi><lb/><hi rendition="#b">durch die</hi><lb/><hi
rendition="#b"><hi rendition="#g">Mark
Brandenburg</hi>.</hi></titlePart></docTitle></titlePage><lb/>
<pb n="[II]" facs="#f0008" />
<pb n="[III]" facs="#f0009" /><titlePage type="main"><titlePart type="main"> <hi
rendition="#b">Wanderungen</hi><lb/>durch die<lb/><hi rendition="#b"><hi rendition="#g">Mark
Brandenburg</hi>.</hi></titlePart><lb/><byline>Von<lb/><docAuthor><hi rendition="#b">Theodor
Fontane.</hi></docAuthor></byline><lb/><docImprint><pubPlace><hi
rendition="#b">Berlin.</hi></pubPlace><lb/><publisher><hi rendition="#g">Verlag von Wilhelm
Hertz</hi>.<lb/>(Be&#x017F;&#x017F;er&#x017F;che
Buchhandlung.)</publisher><lb/><docDate>1862.</docDate></docImprint></titlePage><lb/>
<pb n="[IV]" facs="#f0010" /> <div type="imprint"><p>Osnabrück, Kisling&#x2019;&#x017F;che
Buchdruckerei.</p></div><lb/>
```

(Zitat aus der XML-Datei der „Wanderungen aus der Mark Brandenburg 1“)

[41] Zitat von der Startseite des DTA: http://www.deutschestextarchiv.de/

Dass die Repräsentation des physischen Buches ein Ziel dieser Enkodierung darstellt, zeigt sich an den <pb>-Elementen („pagebreak“) und der Integration von Faksimiles von Buchteilen, zum Beispiel des Einbands, die keinen beziehungsweise keinen enkodierten Text enthalten (Faksimile 1-6). Nach den verschiedenen enkodierten Titelblättern, wird (ab Seite 11) auch das Vorwort, das in der XML-Datei nach dem zitierten Teil folgt, in das <front>-Element integriert. Der Haupttext ist im <body>-Tag enkodiert, das <back>-Tag enthält lediglich die Repräsentation der letzten vier Seiten des Originals, die keinen enkodierten Text umfassen.

Da TEI bislang insbesondere im Rahmen von Digitalisierungsprojekten genutzt wird und kaum für zeitgenössische belletristische Literatur, ist eine Einschätzung hinsichtlich seines Potenzials, auch derartige Texte angemessen enkodieren zu können, schwierig. Über die Gründe für diesen Befund lässt sich nur spekulieren: Urheberrechtliche Bedenken der Verlage gestalten eine Weitergabe digitaler belletristischer Texte auch für wissenschaftliche Zwecke schwierig; Wissenschaftler können die Texte dementsprechend natürlich auch nicht öffentlich zur Verfügung stellen, obwohl Textanalysen auf ihrer Basis auch ohne Veröffentlichung der getaggten Texte (im Sinne von Primärdaten) denkbar wären. Dass Wissenschaftler in Bezug auf aktuelle Belletristik mit einer Enkodierung in TEI zurückhaltend sind, mag auch an der Aufwändigkeit der Enkodierung liegen, die man offenbar eher für bereits kanonisierte Texte in Kauf nimmt. Projekte der Digital Humanities halten sich – zumindest im deutschsprachigen Raum – insofern zumeist an gemeinfreie digitale Texte. Digitale aktuelle Belletristik spielt auch im E-Book-Angebot von Wissenschaftlichen Bibliotheken als Lektüreangebot für Studierende und Wissenschaftlerinnen und Wissenschaftler kaum eine Rolle. Offensichtlich besteht auf Seiten der Studierenden und Wissenschaftler keine entsprechende Nachfrage, was an der bereits beschriebenen Einstellung in Bezug auf digitale Angebote liegen mag, an der eingangs erwähnten kulturellen Prägung dieser Disziplinen, wahrscheinlich jedoch auch daran, dass die reine Lektüre von Primärliteratur am Monitor relativ mühsam ist und sich E-Book-Reader bislang nicht durchgesetzt haben.

Allerdings werden in Einzelfällen mit Verlagen Vereinbarungen getroffen, wie das Beispiel des Digitalen Wörterbuchs der deutschen Sprache des 20. Jahrhunderts (DWDS) zeigt. Ziel ist – dem Vorbild des British National Corpus folgend – die

Erstellung eines Textkorpus, die computerphilologische Erschließung der Texte und die eigentliche Wörterbucherstellung. Im Rahmen der Erstellung des Kernkorpus, das zu circa 26 Prozent aus belletristischen Texten besteht, kam es zur Rechtevereinbarung mit namhaften, vor allem belletristischen Verlagen – darunter dem Aufbau Verlag, der S. Fischer Verlagsgruppe oder auch dem Suhrkamp Verlag.[42]

Die belletristischen Texte aus dem Kernkorpus des DWDS sind – dem Zweck entsprechend – eher zurückhaltend getaggt. Bereits in der „Encoding Description" im TEI-Header der XML-Datei von Maxim Billers Erzählung „Wenn wir einmal reich und tot sind" aus dem 1990 erschienenen gleichnamigen Erzählband (Biller) zeigt sich – als Standard-Voreinstellung – die Zielrichtung der Enkodierung:

```
<samplingDecl>
<p>linebreaks in paragraphs have been omitted (except for poems etc.)</p>
<p>heading lines and footers have been omitted.</p>
<p>blank lines and multiple blank spaces, including paragraph indents, have not been
preserved.</p>
<p>marginal notes have been omitted.</p>
<p>font format (size, type) have not been preserved.</p>
<p>no reproduction of front and back matters.</p>
<p>opener information have not been tagged as such.</p>
<p>preface(s), appendix(ices), index(es) and epilogues (of the author / editor) have
been omitted.</p>
<p>figures and tables have not been preserved.</p>
</samplingDecl>
```

In diesem Fall ist bei der Enkodierung also nicht die physische Gestalt der gedruckten Ausgabe entscheidend – wenn auch die Paginierung mit <pb>-Elementen enkodiert ist –, sondern der Text selbst. Das bedeutet jedoch nicht, dass keinerlei grafische Merkmale enkodiert werden. In diesem Kontext ist folgendes Beispiel aufschlussreich:

```
<p>
<hi rend="italic">Der kämpferische Vorsitzende der »Atlantik-Brücke«, der einst mit seiner
Pistole einen Terroristen in die Flucht geschlagen hatte, bat die Wirtschaftskapitäne,
Mäzene und Senatoren um Ruhe, die Gläser wurden geräuschvoll abgestellt, jemand
benutzte die plötzliche Stille für eine letzte laute Bemerkung, und dann übergab der
Vorsitzende dem Bundespräsidenten endlich das Wort.</hi>
</p>
```

[42] „Bereits frühzeitig hat das Projekt daher Nutzungsvereinbarungen mit Verlagen wie Aufbau, Diogenes Verlag, Eichborn, S. Fischer Verlagsgruppe, Hoffmann & Campe, Kiepenheuer & Witsch, K.G. Saur Verlag, Spiegel, Suhrkamp, Ullstein, ZEIT sowie öffentlichen und privaten Textgebern (z.B. Deutsches Rundfunkarchiv, Digitale Bibliothek) getroffen. Somit kann das Projekt z.B. Werke von Thomas und Heinrich Mann, Martin Walser, Heinrich Böll, Jürgen Habermas oder Victor Klemperer für die Internetrecherchen zur Verfügung stellen." (http://www.dwds.de/textbasis/kerncorpus)

Das highlighting-Element (<hi>) erlaubt die Auszeichnung von typografischen Besonderheiten und Hervorhebungen, wobei das rendition-Attribut (rend=) definiert, welche Hervorhebungsart – in der zitierten Textpassage: kursive Schrift („italic“) – im Originaltext gewählt wurde. Obwohl normalerweise die grafische Gestaltung in der XML-Datei keine Berücksichtigung findet, dokumentiert das gewählte Beispiel, dass einige Gestaltungsaspekte durchaus als semantische Aspekte des Textes verstanden und dementsprechend in der XML-Datei kodiert werden können. Im konkreten Fall wird kursive Schrift ausgezeichnet, die in vielen Texten Billers als semantisches beziehungsweise strukturierendes Element – etwa, wie hier, zur Auszeichnung einer Binnenerzählung – eingesetzt wird und eben keine zu vernachlässigende Größe darstellt.

Wie sich an diesen Beispielen zeigt, wird die Enkodierung der (gedruckten) Originale dem Verwendungszweck angepasst. Von der Auszeichnung einfacher und sofort sichtbarer Strukturmerkmale (wie Abschnitten) bis hin zu komplexeren linguistischen, semantischen oder prosodischen Auszeichnungen wird im digitalen Text im Vergleich zum analogen Text ein philologisches Wissen eingebracht und als Information hinterlegt. Für High-level-Markup können insbesondere digitale Editionen, die reiche Annotationen des Primärtextes (zum Teil auch mit historischem Hintergrund oder intertextueller Anreicherung) bieten, stehen.

Möglich sind also immer mehrere Versionen, die die Ausgabe und somit auch die Interpretation des Textes durch die Rezipienten entscheidend verändern können. Die bibliothekarische, philologische und linguistische Information wandert in den Primärtext. Von einem stabilen Dokument zu sprechen, ist mit Blick auf die unterschiedlichen Schwerpunktsetzungen und den sich daraus ergebenden potenziellen Interpretationen nicht mehr möglich: viel eher ließe sich ein Dokument als Zusammensetzung vieler verschiedener Enkodierungen – zum Beispiel im Sinne der in der 6. These vorgestellten Aggregation – vorstellen. Im Vergleich zum traditionellen Dokument wäre das so entstandene digitale Dokument sehr viel reicher, da es verschiedene Sichten auf einen Text enthalten würde.

Abgesehen von der relativ kleinen Gruppe der Computerphilologen verhält sich die große Masse der Geisteswissenschaftler in Bezug auf digitale Texte jedoch eher zurückhaltend. In diesem Kontext ist die Tradition, der die Computerphilologie entstammt, aufschlussreich. Jannidis unterscheidet drei Phasen der

Computerphilologie: die erste Phase von 1949 bis Mitte der 1980er Jahre ist anfangs insbesondere von Projekten aus der Theologie (wie zum Beispiel der computergestützten Erstellung einer Konkordanz zu den Werken Thomas von Aquins durch Robert Busa) mit Hilfe von Großrechnern geprägt. Die zweite Phase (Mitte der 1980er Jahre bis Ende der 1990er Jahre) ist durch die Verbreitung des Personal Computers und der CD als Speichermedium bestimmt. Die dritte Phase seit Ende der 1990er Jahre steht im Zeichen der Vernetzung, des World Wide Web und ersten Experimenten mit Netzeffekten (vgl. Jannidis 2007). Im angloamerikanischen Raum und im letzten Jahrzehnt auch im deutschsprachigen Raum findet allmählich eine Institutionalisierung der Computerphilologie beziehungsweise – weiter gefasst – der Digital Humanities statt.[43] Inzwischen ist in den USA der Aufgabenbereich der digitalen Texteditionen und ihrer Archivierung, aber auch die Vermittlung von Informationskompetenz für „digital scholars“ teilweise an Universitätsbibliotheken übergegangen.[44] Auch in Deutschland scheint man sich von Unternehmungen des LIS-Bereichs eine verstärkte Nutzung digitaler Möglichkeiten auch in den Geisteswissenschaften zu versprechen. Neben „ideologischen Barrieren“ und „mangelnden technischen Kenntnissen“ kann die zurückhaltende Nutzung digitaler Technologien in der Geisteswissenschaft auch darauf zurückgeführt werden, dass „die bestehenden Techniken in vielen Fällen nicht den geisteswissenschaftlichen Anforderungen entsprechen“ (Burch et al., S. 181). Wünschenswert ist daher eine Mitarbeit von Geisteswissenschaftlern an informatischen Entwicklungen, ein Sich-Einbringen gerade im Bereich der traditionell geisteswissenschaftlich-bibliothekarischen Themen der Strukturierung (digitaler) Texte, der Wissensorganisation im Kontext semantisch basierter Technologien und vor allem eine Offenheit, bislang geschlossene (bibliothekarische und editorische) Räume (OPACs, aber auch TEI-Guidelines) für den Bereich des WWW zu öffnen

[43] Texte und Zentren der Digital Humanties bzw. Humanties Computing werden gegründet, so beispielsweise schon 1976 das Oxford Text Archive von Lou Burnard. Verbände und Zeitschriften entstehen; zu nennen ist insbesondere die britische Association for Literary and Linguistic Computing (ALLC) mit ihrem „Journal of Literary and Linguistic Computing“ und ihre nordamerikanischer Schwesternverband, die Association for Computing in the Humanities (ACH). 1987 gründet Willard McCarty das Diskussionsforum HUMANIST, das er seitdem moderiert. In Deutschland erscheint seit 1999 das „Jahrbuch für Computerphilologie“, das von Georg Braungart, Peter Gendolla und Fotis Jannidis herausgegeben wird.

[44] Vgl. beispielsweise das „Scholars' Lab“ der University of Virginia Library (http://www2.lib.virginia.edu/scholarslab/): „The idea was to combine the resources and expertise of the Library’s successful *Electronic Text* (Etext) and *Geospatial and Statistical Data* (GeoStat) centers with that of UVA’s *Research Computing Support Group* in a physical space that promotes collaboration and experimentation“ (http://www.scholarslab.org/about/).

(beispielsweise auch dadurch, dass WWW-Standards für Regelwerke genutzt werden, anstatt eigene komplexe präskriptive Standards zu entwickeln). In diesem Zusammenhang ist die Forderung der Deutschen Forschungsgemeinschaft (DFG) an „die heute noch weitgehend getrennt operierenden Bibliotheken, Archive und Fachinformationseinrichtungen mit überregionaler Ausstrahlung“ interessant, sich in Zukunft „zu einem kohärenten Gesamtsystem der digitalen Informationsversorgung für die Wissenschaft zu vernetzen“ und „zu Trägern des E-Science-Gedanken“ zu werden (DFG 2006, S. 2). Fraglich ist jedoch, ob mit dem XML-hierarchisch fundierten TEI-Guidelines tatsächlich ein geeignetes System für genuin digitale Texte und das darauf basierende wissenschaftliche Arbeiten gelegt ist. Es wird sich in Zukunft, wenn hoffentlich zunehmend auch genuin digitale Texte mit TEI enkodiert werden, zeigen müssen.

Welche neuen Möglichkeiten ergeben sich für Philologien unter digitalen Bedingungen? In Emulationsszenarien, in denen die digitale Umgebung lediglich den Zugang zu den Quellen erleichtert und die traditionellen Arbeitstechniken unverändert bleiben, werden die Potenziale innovativer Techniken nicht genutzt. Sobald jedoch „digital gedacht“ wird, verändern sich auch die Methoden und Rahmenbedingungen wissenschaftlichen Arbeitens. Beispiele hierfür stellen digitale Bibliotheken, die semantisch basierte Technologien nutzen, und Virtuelle Forschungsumgebungen dar, wobei gilt: Die „line between **portals, digital libraries and VREs** is becoming increasingly blurred“ (Carusi/Reimer, S. 43). In den Geisteswissenschaften gestaltet sich die Entwicklung dieser Plattformen grundsätzlich anders als in den Naturwissenschaften, da mit zwei Herausforderungen umzugehen ist: „heterogeneity of data and processes and low computer literacy of the users“, sodass „an electronic environment that must be extremely flexible, adaptable and extensible but at the same time integrated and easy to use“ das Ziel sein muss (Barbera et al., S. 133).

Semantisch basierte Technologien werden zum Beispiel in den digitalen Bibliotheken genutzt, die im Zuge des zwischen 2006 und 2009 EU-geförderten Discovery Project entstanden sind. Die Absicht des Projekts war, einen Verbund von digitalen Bibliotheken zu bestimmten Themen und Autoren im Bereich der antiken, modernen und zeitgenössischen Philosophiegeschichte zu schaffen.[45] Die technische

[45] Zu nennen ist Presocratics Source (http://presocratics.daphnet.org/), Socratics Source (http://socratics.daphnet.org/), Diogenes Laertius Source (http://laertius.daphnet.org/), Modern

Infrastruktur stellt Talia, ein Open-Source-System für semantische digitale Bibliotheken, das speziell für geisteswissenschaftlichen Inhalt entwickelt wurde und Interoperabilität mit anderen semantisch basierten Tools und Anwendungen ermöglicht. So ist die Talia Knowledge Base ein RDF Triplestore, zusätzlich wird RDFS und OWL zum Aufbau von Ontologien genutzt, die die Community selbst entwickeln soll. Grundsätzlich wird zwischen Primär- und Sekundärquellen unterschieden, die Discovery-Projekte sind stets sowohl digitale Bibliotheken als auch Open-Access-Publishers. Weitere Tools ermöglichen den Wissenschaftlern die semantische Anreicherung des Materials durch Annotationen, den Ausbau von RDF-Verbindungen und das Weiterschreiben von Ontologien. Eine interessante Komponente der Discovery Federation ist, dass sie als „peer-to-peer network of interconnected libraries“ konzipiert sind (Barbera et al., S. 133), die bidirektionale semantische Links zwischen den digitalen Objekten unterschiedlicher Bibliotheken erlauben und so zwar die Verwaltung der eigenen Objekte den jeweiligen Bibliotheken überlassen, jedoch die Anreicherung von Material aus anderen Bibliotheken, also virtuelle Sammlungen ermöglichen. Insofern können die Discovery-Projekte auch als eine große verteilte digitale Bibliothek für Primär- und Sekundärquellen verstanden werden, die der üblichen Fragmentierung in kleine digitale Bibliotheken ohne Verbindung untereinander ein Alternativmodell entgegenstellt.

Als ein Beispiel kann Nietzsche Source dienen. Zusätzlich zur synoptischen Edition der Faksimile Gesamtausgabe und der Kritischen Gesamtausgabe – in TEI kodiert – mit einer Vielzahl von Funktionalitäten für die Präsentation der Inhalte wurde semantisch basierte Technologie integriert: beispielsweise werden die Metadaten der Manuskripte in RDF übertragen und eine Ontologie (in RDFS) zur formalisierten Beschreibung von Handschriften verwendet, so dass etwa automatisiert ein Stemma erstellt werden kann, das natürlich noch der menschlichen Überprüfung bedarf, jedoch computergenerierte Hypothesen über Abhängigkeiten zwischen Textfassungen liefert.

Das Anwachsen des digitalen Materials, bei dem zunehmend auf Interoperabilität geachtet wird, stellt durch das ihm innewohnende Potenzial zur flexiblen Aggregation großer digitaler Korpora die Bedingung für den „Paradigmenwechsel vom Lesen zum

Philosophy Source (http://modernsource.daphnet.org/), Wittgenstein Source (http://www.wittgensteinsource.org/) und Nietzsche Source (http://www.nietzschesource.org/).

Rechnen der Literatur“ dar. Textmining-Verfahren ermöglichen Methoden des „distant readings“, die die bestehende Philologie auch insofern ändert, als dass sie Texte jenseits des Kanons in den Blick nehmen kann (vgl. Lauer).[46]

Um solche Möglichkeiten nutzen zu können und Wissenschaftler von den mechanischen Tätigkeiten zu entlasten, fördert das Bundesministerium für Bildung und Forschung (BMBF) im Rahmen der D-Grid-Initiative das Projekt TextGrid, eine – auf einer Grid-Infrastruktur basierende – Virtuelle Forschungsumgebung für Geisteswissenschaftler zur verteilten und kooperativen wissenschaftlichen Textverarbeitung. Die Basis bildet das TextGrid-Repositorium, das auch zur Langzeitarchivierung der digitalen Texte dienen soll. Außerdem bietet das TextGridLab digitale Werkzeuge wie einen XML-Editor, Lemmatizer, Tokenizer, Metadaten-Annotator, ein Wörterbuch-Tool und Instrumente für das Projektmanagement.[47] Neben den beteiligten Wissenschaftlern spielen Wissenschaftliche Bibliotheken eine Schlüsselrolle beim Aufbau von TextGrid, wobei die Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB) als Koordinatorin wirkt. Gerade für „die Vernetzung und Abstimmung zum Beispiel über Standards und technische Schnittstellen, auch über Disziplingrenzen hinweg“ ist es „unabdingbar, dass Wissenschaftler verschiedener Fachbereiche und Vertreter informationswissenschaftlicher Disziplinen die Forschungsumgebungen gemeinsam entwickeln“ (Neuroth et al., S. 163). Ziel des Vorhabens ist es, dass Wissenschaftler auf Inhalte und Forschungsdaten zugreifen können und ihre Arbeitsprozesse in der Virtuellen Forschungsumgebung durch die bereitgehaltenen Tools unterstützt werden. Analog zu den naturwissenschaftlichen Primärdaten können die Geisteswissenschaften beispielsweise Quellenkorpora, Annotationen oder Sammlungen von global verteilten Objekten, die für aktuelle Forschungsfragen relevant sind, zur Verfügung stellen, nachnutzbar machen und so die Datengrundlage für einzelne Forschungsfragen verbessern: Bild- und Textdigitalisate, dazugehörige Metadaten oder auch fachwissenschaftlich angereicherte Daten, wobei auf Interoperabilität zu achten ist. Als semantisches Best-practice-Beispiel führt TextGrid TEI an (vgl. Neuroth et al.). In allen Bereichen wird mit Anwachsen der Repositorien und der zu verwaltenden Daten die Frage nach Langzeitarchivierung und Verfügbarhaltung und vor allem nach der Präsentation und Auffindbarkeit der

[46] Vgl. dazu auch den einflussreichen Beitrag von Gregory Crane (Crane 2006).
[47] Eine Betaversion kann bereits heruntergeladen werden: http://www.textgrid.de/beta.html

digitalen Objekte immer wichtiger werden. Wie die Europeana experimentieren – der „Landscape Study“ des JISC zufolge (Carusi/Reimer) – auch Virtuelle Forschungsumgebungen international mit semantisch basierten Technologien. Der Nutzen der Mitarbeit an VREs für Bibliotheken ist dabei vielfältig und wird von Carusi/Reimer wie folgt zusammengefasst:

> Research-oriented **libraries get more involved with VREs**, for a variety of reasons: 1) VREs are seen as a good way to feature library resources, especially domain specific ones that researchers are not always aware of; 2) The more libraries are concerned with curating digital research output, the more they realise that the best way to make that task easier is to be involved in the creation of these materials from the start; 3) Libraries who see themselves as supporting the whole research lifecycle want to better understand the impact of VREs on that process. As libraries have always been places of communication and collaboration it may be seen as a logical consequence that **digital libraries become more like VREs**. (Carusi/Reimer, S. 44)

Bibliotheken können beim Aufbau digitaler (Geistes-)Wissenschaften ihre traditionellen Stärken im Bereich von Metadaten, Wissensorganisation und Archivierung einbringen, müssen jedoch gleichzeitig auch ihre Kompetenzen im Bereich digitaler Technologien weiter ausbauen und vor allem in Dialog mit anderen Institutionen und Berufsfeldern treten. Wünschenswert wäre nicht nur eine Kommunikation zwischen Wissenschaftlern und Informatikern, sondern auch mit belletristischen Verlagen über Standards, Formate, neue Dokumentformen (vgl. These 6) und Kooperationsmöglichkeiten.

Zusammenfassend ist zu konstatieren, dass sich unter digitalen Bedingungen der Wissenschaftsprozess in all seinen Teilbereichen verändert: Neuartige Forschungsmethoden und -werkzeuge entstehen, die Rahmenbedingungen geisteswissenschaftlicher Forschung wandeln sich durch die Förderung kooperativer Wissenschaft – beispielsweise im Kontext Virtueller Forschungsumgebungen und Virtueller Fachbibliotheken –, neue Publikations- und Kommunikationsformen bilden sich aus. Es ist auch in den Geisteswissenschaften langfristig mit einer Zunahme der Publikation von Primärdaten zu rechnen, die Bedeutung digitaler Publikationen insgesamt und im Open Access wird zunehmen.[48] Das bedeutet zugleich einen

[48] Einen im deutschsprachigen Raum neuartigen Weg schlägt Berlin Academic ein: „Auf einer neu errichteten Online-Plattform veröffentlicht Berlin Academic sein Programm nach Open-Access-Prinzipien und unter Creative-Commons-Lizenzen. Gleichzeitig werden sämtliche Titel über Print on Demand sowie in verschiedenen E-Book-Formaten vertrieben. Berlin Academic veröffentlicht auf Deutsch und Englisch. Thematische Schwerpunkte sind geistes- und sozialwissenschaftliche Disziplinen, insbesondere Zeitgeschichte (Holocaust- und Totalitarismusforschung), Soziologie (Migration, Urbanistik, Arbeitswelten), Politikwissenschaften (Globalisierung, Global Governance), Philosophie (speziell im Spannungsfeld mit Natur- und Neurowissenschaften) sowie Kulturwissenschaften und Linguistik“ (http://www.buchmarkt.de/content/41758-berlin-verlag-gruendet-berlin-academic.htm). Im Gegensatz zum elektronischen Publizieren in universitären Repositorien blieben

erhöhten Service- und Schulungsbedarf auf Seiten der Geisteswissenschaftler, dem Wissenschaftliche Bibliotheken in Kooperation mit ihren Rechenzentren entgegenkommen sollten. Zudem resultiert aus der Zunahme digitaler Forschungsdaten eine Nachfrage nach nachhaltigem Daten- und Informationsmanagements und nach leistungsfähigen digitalen Infrastrukturen. Es ist also anzunehmen, dass Wissenschaft und Bibliotheken in Zukunft noch enger zusammenarbeiten werden.[49]

die Vorteile des Publizierens in einem Verlag insofern bestehen, als dass die Reputation durch die Aufnahme ins Programm von Berlin Academic und ein Marketing für die Publikation gewährleistet würden – ein Modell, das der Kultur der Geisteswissenschaften sicherlich entgegenkommt.

[49] Die DFG fordert dementsprechend – beispielsweise im Kontext der Sonderforschungsbereiche – „die systematische Zusammenarbeit des Sonderforschungsbereichs mit Informationseinrichtungen“. Der Antrag für „Service-Teilprojekte zur Informationsinfrastruktur“ sollte „von einer Wissenschaftlerin oder einem Wissenschaftler im SFB und gleichzeitig von einer weiteren Person, die im Bereich der Informationsinfrastruktur verantwortlich beschäftigt ist (z.B. aus der Universitätsbibliothek oder dem Universitätsrechenzentrum an der antragstellenden Hochschule)“ eingereicht werden (DFG 2009).

4 Digitale Belletristik als allgemeines Kulturgut in Öffentlichen Bibliotheken

E-Books und E-Journals gehören im wissenschaftlichen Bereich bereits zum Standard-Angebot von Bibliotheken. Bei elektronischen Publikationen von Publikumsverlagen – insbesondere im Bereich Belletristik – ist eine derartige Entwicklung noch nicht zu beobachten. Ein Grund ist in der traditionell sehr viel engeren Verknüpfung der Rezeption von Schöner Literatur mit dem physischen Buch zu suchen, insbesondere wenn die Lektüre nicht wissenschaftlichen Zwecken dient. Darüber hinaus scheinen gerade belletristische Verlage große Bedenken zu haben, dass die Urheberrechte unter digitalen Bedingungen ausgehöhlt werden könnten. Obwohl die Förderung von Open Access wissenschaftliche Publikationen betrifft und wiederum nichts mit Google Books zu tun hat, unterzeichnete 2009 eine große Anzahl von Schriftstellerinnen und Schriftstellern und eine Reihe von belletristischen Verlagen den Heidelberger Appell,[50] der als Ausdruck der Verunsicherung in Zeiten eines Wandels der Kulturtechniken gelesen werden kann. Verursacht durch das Urheberrecht ist unter digitalen Bedingungen auch in Bezug auf belletristische Literatur eine Angebotslücke zwischen älteren und aktuellen Publikationen entstanden. Im WWW ist bereits eine Auswahl gemeinfreier Werke über Google Books,[51] das Projekt Gutenberg,[52] zeno.org[53] oder in Zukunft über die Deutsche Digitale Bibliothek (DDB)[54] beziehungsweise Europeana[55] verfügbar. Gegenwartsliteratur, die seit Mitte des letzten Jahrhunderts entstanden ist, ist dagegen bislang in digitaler Form kaum zugänglich.[56] Es wäre wünschenswert, dass die DDB und Europeana für verwaiste Werke eine Regelung mit den zuständigen Verwertungsgesellschaften findet beziehungsweise die Verhandlungen mit Verlagen zu dem Ergebnis führen, dass ein „'One Stop Shop', wo der Nutzer das Werk, das ihn interessiert, auch gleich elektronisch ausleihen oder kaufen kann“, entsteht.[57] Für

50 http://www.textkritik.de/urheberrecht/index.htm
51 http://books.google.de/
52 http://gutenberg.spiegel.de/
53 http://zeno.org
54 http://www.deutsche-digitale-bibliothek.de/
55 http://www.europeana.org/portal/
56 Eine Digitalisierung der Backlist von belletristischen Verlagen wäre – ähnlich dem Angebot von de Gruyer (http://www.degruyter.de/cont/glob/ebooks.cfm) – begrüßenswert.
57 Vgl. Dworschak 2010, Zitat S. 144.

aktuelle Titel zeichnen sich in Deutschland inzwischen erste Ansätze eines E-Book-Marktes ab: mit libreka! hat der Börsenverein des Deutschen Buchhandels mithilfe einer Tochterfirma – recht erfolglos – erste Schritte gewagt;[58] Dienstleister stellen für Verlage E-Books her und optimieren sie beispielsweise auch für Smartphones.[59]

In Wissenschaftlichen Bibliotheken und im Bereich des akademischen elektronischen Publizierens sind bislang XML und PDF die bevorzugten Formate für linearen Text.[60] Offensichtlich ist vor allem PDF ohne Digital Rights Management (DRM) mit der Möglichkeit eines Remote Access und der Option zur Anfertigung von Papierausdrucken dasjenige Format, das in der Wissenschaft und Bibliotheken derzeit bevorzugt wird. So ergibt die Umfrage unter wissenschaftlichen Bibliothekarinnen und Bibliothekaren, die die an der Stanford University angesiedelten HighWire Press durchführte:

> Forty-one percent (53) of participants indicated that users prefer ebooks in PDF format. This strong preference for PDF is not surprising. Ebooks are new and most users have little experience with them in formats other than PDF. Only 12% of participants (16) indicated that HTML is the preferred format. This is a little surprising given that the standard for journals, both commercial and non-commercial, is to offer articles in both HTML and PDF. [...]
>
> Only four participants (3%) indicated that users prefer ebooks optimized for dedicated ebook devices or other mobile devices. Two others selected “Other” and mentioned mobile devices. These results indicate librarians believe that there is a very low demand for ebooks optimized for ebook or mobile devices. However, the market for mobile ebook readers is changing rapidly and as one participant stated, “mobile device access may soon be the most important factor for most users.” (2009 Librarian eBook Survey, S. 26)

Die Vorliebe für PDF geht mit der zweiten These konform: In der Übergangsphase vom analogen zum digitalen Paradigma werden traditionelle Kulturtechniken im Digitalen imitiert. Tatsächlich lassen sich PDF-Dokumente relativ leicht in analoge Denkweisen einpassen, bilden sie doch stabile und nicht-veränderbare Dokumente ab.

Im Unterschied zu Wissenschaftlichen Bibliotheken konzentrieren sich Öffentliche Bibliotheken im Bereich E-Books stärker auf Sachbuch- und in letzter Zeit auch auf belletristische Literatur. Es mag an Titelangebot und Publikum liegen, dass sich im

---

[58] Erfolglos insbesondere wegen des umständlichen Kopierschutzes der E-Books und der damit nutzungsunfreundlichen Funktionalitäten. Zudem sind die meisten deutschsprachigen Verlage bislang mit der Lieferung von E-Book-Titeln noch zu zurückhaltend. Abgesehen von libreka! bestehen mit www.bol.de, www.libri.de oder www.ciando.de weiter Plattformen zum Verkauf von E-Books.

[59] Z. B. Kreutzfeld digital (http://www.kreutzfeldt.de/), pagina (http://www.pagina-online.de/) oder Satzweiss.com (http://satzweiss.com/), die auch eine E-Book-Applikationen auf facebook anbieten (http://www.buchreport.de/pressemitteilungen/pressemitteilungen/datum/2010/03/02/readboox-die-erste-e-book-applikation-auf-facebook.htm); textunes stellt iPhone-Apps zum Lesen von E-Books her (http://www.textunes.de).

[60] Vgl. http://ebookstandards.ub.uni-muenchen.de/wiki/Hauptseite

Bereich Öffentlicher Bibliotheken und auf dem E-Book-Markt der Publikumsverlage ein anderes Format durchsetzt: EPUB. Seit 2007 ist EPUB (**E**lectronic **PUB**lication) der offizielle Standard des International Digital Publishing Forum (IDPF), der Handels- und Standardisierungsorganisation für die Entwicklung und Verbreitung digitaler Publikationen.[61] Neben EPUB existieren – vor allem auf dem US-amerikanischen E-Book-Markt – noch zahlreiche freie und prorietäre Formate, die zum Lesen von Text auf mobilen Endgeräten in Frage kommen.[62] Zumindest in Deutschland scheint die Format-Standardisierung jedoch die gewünschte Wirkung zu haben: Publikumsverlage bieten ihre Titel fast ausschließlich im EPUB-Format (oder als PDF) an.[63] In diesem Kontext ist zu betonen, dass für Verlage – und in Zukunft auch für Bibliotheken mit Pflichtexemplarrecht – nicht das Format EPUB, das auch durch Konvertierung von PDFs hergestellt werden kann, sondern ein Umstieg auf XML-Systeme der gravierende Einschnitt ist: Sei es – für die Verlage – als Voraussetzung für Cross-Publishing, sei es – für Bibliotheken – als Vorstufe für Archivierungsformate. Beispielsweise hat Pagina, ein Dienstleister für Verlage im Bereich des XML-Publishing, zusammen mit der Holtzbrinck-Verlagsgruppe einen XML-Workflow und auch eine DTD (Document Type Definition) entwickelt, deren Lizenz in Zukunft auch mittlere und kleinere Verlage erwerben könnten.[64] Im Unterschied zum PDF-Format, in dem der Text in der Anzeige nicht umbrochen, sondern starr – wie im Buchsatz – dargestellt wird, sind EPUB-basierte E-Books „reflowable“, das heißt, der Text passt sich automatisch an die Bildschirmgröße des jeweiligen Endgeräts an. Texte im EPUB-Format ermöglichen damit nicht nur die Lektüre – beispielsweise in Adobe Digital Editions – am PC, sondern auch ein Lesen auf dem Smartphone oder anderen mobilen Endgeräten.

Kann eine EPUB-Datei analog zum Buch als Dokument verstanden werden? Um die Unterschiede zwischen digitalem Text in EPUB- und Buchformat – in der Ausgabe und der Produktion – zu verstehen, ist der Aufbau einer EPUB-Datei aufschlussreich. Der EPUB-Standard umfasst drei offene Standards: die Open Publication Structure (OPS) für die Festlegung des Formats der eigentlichen Inhaltsdateien, das Open

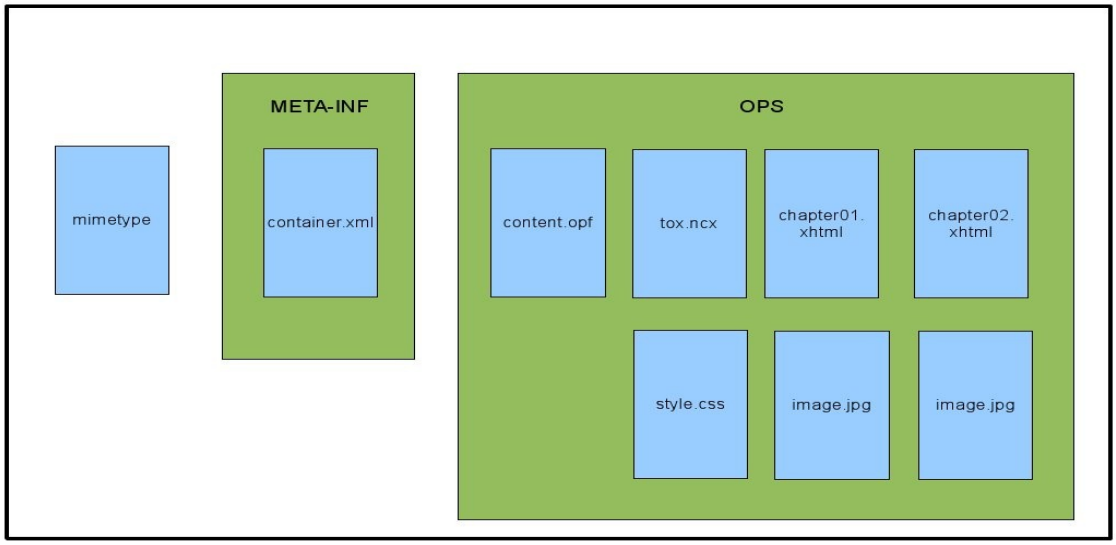
[61] Vgl. http://www.openebook.org/specs.htm

[62] Zumindest werden die E-Books im iBooks-Store, den Apple derzeit plant, im EPUB-Format vorliegen. Auf Amazons relativ weit verbreitetem Kindle werden jedoch beispielsweise nur E-Books im hauseigenen proprietären AZW-Format angezeigt. Andere Formate sind z. B. BB EBook (Sony) oder MOBI/PRC (Mobipocket; lesbar auf iLiad, Cybook und vielen PDAx).

[63] Vgl. zum Beispiel www.libreka.de

[64] Nach Angaben von Pagina ist TEI aufgrund der Komplexität nicht als Auszeichnungssprache in Erwägung gezogen worden.

Packaging Format (OPF) für den Aufbau und die Kompression der EPUB-Datei, wobei bei der Adressierung von Metadaten Dublin Core verwendet wird, und das Open Container Format (OCF), das als ZIP-basierter Standard festlegt, in welcher Ordnerstruktur die Daten abgelegt beziehungsweise nach dem Download der EPUB-Datei zu entpacken sind (vgl. z. B. Roesler-Graichen/Schild, S. 23ff.).



Im Kontext des Themas Dokument sind vor allem folgende Charakteristika einer EPUB-Datei von Interesse: die Trennung von Inhalt (Text) und Repräsentation (Stylesheet) beziehungsweise von Struktur (XHTML) und Layout (CSS), die Trennung von Text und Bild und schließlich die Navigationsmöglichkeiten (NCX).

Der Textinhalt befindet sich in einer EPUB-Datei bei längeren Publikationen, die eine Kapitelstruktur aufweisen, zumeist in unterschiedlichen XHTML-Dateien im OPS-Verzeichnis. Grob lassen sich in einer solchen XHMTL-Datei Header und Textkörper voneinander unterscheiden: Während im Header Informationen über Zeichensatz (inzwischen meist Unicode), Standard, das hineinzuladende Stylesheet und Ähnliches enthalten sind, befindet sich der angezeigte Text innerhalb des „body“-Tags. Die Auszeichnung in belletristischen Texten wird von der Herstellung der Publikumsverlage relativ minimalistisch gehalten. In XHTML wird nicht mehr typografisch strukturiert – obwohl die Ausgabe mittels Stylesheet zurzeit größtenteils der einzige Nutzen dieser Tags ist –, sondern nach strukturellen Kriterien. Angegeben werden also semantisch-strukturelle Funktionen, zum Beispiel bestimmte Klassen von Sinnabschnitten (die allerdings meist mit typografischen Konventionen assoziiert

werden). Diese Gliederung in Sinnabschnitte basiert auf Entscheidungen des Autors oder der Autorin beziehungsweise denen des Lektorats, wobei in Romanen einige Konventionen – beispielsweise in der Wiedergabe von Dialogpassagen – existieren, die sich im Digitalen anscheinend nicht verändert haben und die somit meist auch von Herstellung, Setzereien und/oder Dienstleistern vollzogen werden können. Der nach narrativen Kriterien leicht strukturierte Inhalt ist also zunächst unabhängig von der Repräsentation und damit potenziell auch von Maschinen für Textmining-Verfahren verwertbar. Dass selbst das strukturelle Markup einer hermeneutischen Tätigkeit in der Produktion entstammen muss, stützt die im ersten Kapitel formulierte dritte These zur Diskretheit. Denn alle digitalen Dokumente und damit also Texte, die potenziell auch von Menschen und nicht nur von Maschinen gelesen werden sollen, erfordern hermeneutische Tätigkeit – und zwar bereits in ihrer Produktion. Nicht nur semantisches Markup, auch strukturelles Markup basiert auf kontextabhängigen Entscheidungen, die die Semantik des Textes beeinflussen. Die reine Binärkodierung kann keinen adäquaten Textbegriff und erst recht keinen Dokumentbegriff stützen (vgl. hierzu auch die fünfte These), der stets das unter anderem durch Markup hinzugefügte kulturell überlieferte Wissen im Kontext hermeneutischer Aktivität – sei es von Autorinnen und Autoren, Lektorat, Herstellung oder Setzerei – erfordert.

Interessant ist eine weitere Auszeichnung, die im gewählten Beispiel auftaucht und zurzeit noch sehr verbreitet in Texten im EPUB-Format ist: Mittels identifier-Attribut wird die Paginierung des gedruckten Buches eingefügt. Aufgrund des Wegfalls von „Seiten“ (zum Beispiel der Schmutztitelseite) im Digitalen wirkt die Paginierung besonders absurd. Zum einen zeigt sich daran, dass es sich bei heutigen EPUB-Dateien häufig um rekonvertierte digitale Druckdaten handelt, beziehungsweise dass es, selbst wenn es bereits in einem XML-Workflow entstandene Dateien sind, ein Druckdokument gibt, das als Referenz herangezogen wird; zum anderen demonstriert es, dass es bislang noch keine adäquaten Referenzierungsmöglichkeiten auf Mikrostrukturen innerhalb eines digitalen Textes gibt.[65] Obwohl EPUB als Format anders als PDF nicht mehr auf starre Druckseiten-Imitate zurückgreift, wird durch die kleinen Seitenzahlen, die sich in der Ausgabe meist am rechten Rand der Anzeige befinden, auf eine Funktionalität aus der

[65] Eine genauere Referenzierung auf Textstellen – beispielsweise durch Zeichen- statt Seitenzählung – ist ein Beispiel für eine neue bzw. bessere Funktionalität im Digitalen, die allerdings bislang – aufgrund der Emulation des Druckmodus – noch nicht realisiert worden ist.

Druckwelt zurückgegriffen: ein Atavismus im digitalen Text, der auf Kodierungsebene aus der Baumstruktur ausbricht und auf Anzeigeebene der Repräsentation des Textes, der eher an eine Schriftrolle als einen Kodex erinnert, zuwiderläuft. Anders als im PDF-Format ist die Leserichtung im EPUB-Format durch Herunterscrollen bestimmt, wobei es einfache String-Suchfunktionen wiederum ermöglichen, mit der linearen Lesart zu brechen.

Die Form des Textes kommt erst mit dem Stylesheet ins Spiel, das im Prolog der XHMTL-Datei gekennzeichnet, zur Ausgabe des Textes in die XHTML-Datei hineingeladen wird und die strukturelle Auszeichnung nutzt. Auf dieser Trennung von Form und Inhalt basiert Cross-Media-Publishing, das ohne großen Aufwand eine unterschiedliche Repräsentation von Text mittels unterschiedlicher Stylesheets ermöglicht. Das bedeutet jedoch auch, dass das stabile Dokument nicht mehr die Inhalt und Form integrierende Anzeige auf einem Endgerät sein kann. Und es spricht für das unter der fünften These vorgestellte Textmodell, demzufolge Text eine Ordered Hierarchy of Content Objects ist, deren Repräsentation nicht essentiell ist. Allerdings ist dieses Modell zu hinterfragen, wenn an die bereits erwähnten Konventionen gedacht wird, die sich für die Repräsentation von narrativen Strukturen weiterzutradieren scheinen. Das folgende Beispiel aus dem CSS-Stylesheet kann diesen Umstand demonstrieren:

```
p.para
{       margin-top:1em;
        margin-bottom:0em;
        text-align:justify;
        margin-right:0em;
}
 [...]
p.indent-para
{       margin-top:0em;
        margin-bottom:0em;
        text-align:justify;
        text-indent:1.0em;
        margin-right:0em;
}
```

(Zitat aus dem Stylesheet der EPUB-Datei des Romans von Dara Horn (Horn 2009).)

Die Selektoren (in diesem Fall „p.para“ und „p.indent-para“) sorgen dafür, dass die Elemente in der XHTML-Datei gefunden werden. Daraufhin kann die Formatierung, die unter jedem Selektor im Stylesheet exakt angegeben ist, auf die entsprechenden Elemente angewendet werden. Da die Anzeige sich je nach Gerät verändert, werden

keine absoluten Maßeinheiten verwendet: „em“ entspricht der Schriftgröße (der Entfernung vom höchsten Punkt der Oberlänge bis zum tiefsten Punkt der Unterlänge der Buchstaben einer Schrift) und ist damit eine relative Maßeinheit.

Es ist anzunehmen, dass sich die Herstellungs- und Gattungskonventionen in Bezug auf digitale Texte tradieren werden oder sich in anderer Weise ausbilden. Die Konventionen sind nun als Regelkatalog im Stylesheet formuliert. Natürlich sind die Stylesheets veränderbar; dieser Umstand macht jedoch im Vergleich zum gedruckten Text, der beispielsweise auch zum Zitieren in anderen Schriften abgetippt werden kann, keinen entscheidenden Unterschied. Das kulturelle Wissen wird – nach einer anfänglichen Verunsicherung – im digitalen Text ebenso aufgehoben wie im physischen Text und zusätzlich sogar noch im Stylesheet explizit gemacht.

Derzeit werden die mächtigeren XSL- und XSLT-Stylesheets für EPUB-Dateien, die belletristische Texte enthalten, kaum genutzt. Um die Funktionalitäten dieser Stylesheets, wie beispielsweise die Veränderung der Reihenfolge der Sortierung von Daten in der Ausgabe, nutzen zu können, müsste allerdings bedeutend mehr ausgezeichnet werden, ohne dass es im linearen Textformat bislang einen speziellen Nutzen zeigt. Für experimentelle digitale Literatur mögen XSL und XSLT jedoch in Zukunft passendere Transformationssprachen sein.

Die dynamische Anzeige von EPUB-Dateien hat einige Auswirkungen auf Herstellungsstandards. Traditionell wird in Deutschland insbesondere in Bezug auf belletristische Literatur viel Wert auf die grafische Gestaltung von Büchern gelegt. Durch die dynamische Anpassung bei der Ausgabe ist es nicht mehr möglich, Texte in gewohnter Form zu setzen. Bislang sind Cover und Paratexte selten Teil der EPUB-Datei,[66] sondern werden lediglich für Verkaufs- und Bibliothekplattformen mitgeliefert, ein Umstand, der den technischen Möglichkeiten der Integration von Bild und Text im Digitalen eigentlich zuwiderläuft – tatsächlich werden häufig in der XHMTL-Datei mittels Verlinkungen auf die Quelle im Container der EPUB-Datei mitzuliefernde Bilder beim Anzeigen in den Text hineingeladen. Das Fehlen des Covers kann als weiterer Hinweis auf das Durchlässigwerden der Dokumentgrenzen verstanden werden, wobei für gewöhnlich in der Anzeige der EPUB-Datei nicht auf Coverdarstellungen im WWW verlinkt wird, diese Information ist überhaupt nicht in

[66] Dieser Umstand erklärt sich zum Teil wahrscheinlich auch aus bildrechtlichen und gestaltungsrechtlichen Gründen – die entsprechenden Vereinbarungen zur Verwendung der Cover in EPUB-Dateien sind möglicherweise noch nicht Teil der Verträge mit Grafikern geworden.

der Datei abgelegt.

Greift das EPUB-Format einerseits durch die Leserichtung Charakteristika der Schriftrolle auf, lässt sich diese Analogie durch zwei genuin digitale Funktionalitäten nicht halten: die Volltextsuche und die Navigation im Inhaltsverzeichnis mittels der NCX-Datei (Navigation Control file for XML), wobei beides keine EPUB-spezifischen Funktionalitäten sind, sondern auch schon für Texte im PDF-Format realisiert werden können. Die Navigation Map der NCX-Datei sieht folgendermaßen aus:

```
<navMap>
                    <navPoint id="title" playOrder="1">
        <navLabel>
                    <text>Titel</text>
                    </navLabel>
                        <content src="xhtml/titlepage.html"/>
                        </navPoint>
                    <navPoint id="dedication" playOrder="2">
        <navLabel>
                    <text>Widmung</text>
        </navLabel>
                        <content src="xhtml/dedication.html"/>
[...]
</navMap>
```

(Zitat aus der NCX-Datei der EPUB-Datei des Romans von Dara Horn (Horn 2009).)

Die Datei enthält ein detailliertes, häufig auch ein verschachteltes Inhaltsverzeichnis. Im Attribut „playOrder“ wird einem Eintrag eine Position innerhalb der Reihenfolge der Texte zugeordnet. Letztlich folgt die Navigation Map der Funktionsweise von Inhaltsverzeichnissen in Büchern beziehungsweise holt diese Funktionalität in die EPUB-Datei hinein, denn Binnenstrukturierung und nicht-lineares Lesen, also Blättern und Suchen, ermöglichte erst der Kodex. Greift das EPUB-Format in seiner Anzeige auf Charakteristika der Schriftrolle zurück, muss die Funktionalität des Springens im Text durch genuin digitale Funktionsweisen aufgehoben werden, was mittels der NCX-Datei geschieht.

Insgesamt besitzt das EPUB-Format – ähnlich wie PDF – jedoch wenig genuin digitale Funktionalitäten, ein Umstand, der vor allem damit zu tun hat, dass EPUB-Dateien – unter anderem aus Gründen des zu schützenden Urheberrechts, das sich im Bereich belletristischer Literatur grundsätzlich anders darstellt als im Bereich wissenschaftlicher Literatur – frenetisch abgeschottet und bislang kaum in vernetzter digitaler Umgebung angeboten werden. Der Kauf beziehungsweise die Distribution

findet unter vernetzten Bedingungen statt, die Rezeption jedoch potenziell nicht – etwa aufgrund von auf dem deutschen Markt erhältlichen E-Book-Readern, die bislang zum größten Teil noch nicht internetfähig sind.[67] Dadurch werden buchanaloge Entitäten geschaffen, die den Dokumentbegriff aufrechterhalten, anstatt ihn neu zu denken. Vor allem das Digital Rights Management (DRM), mit dem EPUB-Dateien auf verschiedenartige Weise versehen werden können, hat in den letzten Jahren Debatten entfacht. Dabei wird zumeist ein Umstand, der das EPUB-Dokument ohne DRM vom traditionellen Dokument grundlegend unterscheidet, gar nicht diskutiert:

> Bei ungeschützten EPUB-Dateien handelt es sich faktisch um „offene“ (X)HTML-Dateien. Ob nun EPUB-Dateien illegal verbreitet werden oder nicht, das sei an dieser Stelle einmal vernachlässigt. Vielmehr geht es darum festzustellen, dass die Integrität der Inhalte bei ungeschützten EPUB-Dateien nicht gewährleistet ist. Das heißt konkret, dass Veränderungen der Texte an sich prinzipiell möglich sind, man muss die entsprechende EPUB-Datei dazu lediglich entzippen – eine Horrorvorstellung für Autoren und Verleger gleichermaßen.
>
> Die Entwicklung von digitalen Wasserzeichen und Hashwert-Berechnungen für die neuen Formate liegt in weiter Ferne. (Posth, S. 61)

Dem kann noch hinzugefügt werden, dass jene Verlage, die besonders anspruchsvoll gestaltete Produkte liefern möchten und eine ungeschützte Datei, die spezielle Schrifttypen in den Text einlädt, offen anbieten, auch hinsichtlich der meist ebenfalls urheberrechtlich geschützten Schriften in juristische Schwierigkeiten geraten können. Im Kontext des Dokumentbegriffs ist die Veränderbarkeit nicht nur der Form, sondern auch der Inhalte, ohne dass eine solche „Fälschung“ besonders kompliziert oder auffällig wäre, ein Angriff auf die mit dem Dokumentbegriff verbundene Integrität.

Dass das EPUB-Format – wenn auch in geringerem Maße als PDF – eine Emulation analoger Dokumente im Digitalen darstellt, zeigt sich vor allem daran, dass die potenzielle Vernetzung von Ressourcen nicht mit einbezogen wird und an einer Entität, einer mit bestimmten Schutzmechanismen versehenen Datei festgehalten wird, wenn auch die Grenzen, wie gezeigt wurde, bereits durchlässig geworden sind.

Diese Diagnose wird durch die Art, wie E-Books in Öffentlichen Bibliotheken angeboten werden, weiter bekräftigt. Verständlicherweise möchte nicht jede – kleinere – Öffentliche Bibliothek in Verhandlungen mit Verlagen treten und häufig besteht auch gar nicht die personelle Kompetenz und Kapazität, um die Daten zu hosten und anzubieten. Die ekz hat daher 2005 die Tochterfirma DiViBib gegründet,

[67] Erste Schritte in Richtung Vernetzung vollzieht z. B. der txtr Reader (http://reader.txtr.com/de/start/txtr/).

die als Aggregator für E-Book-Rechte und -Daten auftritt, die Daten hostet und eine in den Webauftritt der jeweiligen Bibliothek integrierbare Plattform zur Präsentation des Angebots an elektronischen Büchern bietet. So erklärte 2005 denn auch der Geschäftsführer der ekz: das „Geschäftsmodell öffentlicher Bibliotheken in die digitale Welt und auf das Internet zu transformieren – das ist die Idee, die hinter DiViBib steckt“.[68] Dementsprechend funktioniert das Geschäftsmodell, ein Emulationsszenario, wobei die Gründe für die Entscheidung für ein derartiges Verfahren nicht nur bei der gestaltenden Firma, sondern auch bei den Verlagen zu suchen sind und der Impuls, also einen Aggregator für digitale Inhalte und Dienstleistungen zu schaffen, grundsätzlich begrüßenswert ist. Sprechend ist der Name für dieses Angebot: „Onleihe“. Analog zu Büchern werden hier E-Books auf einem „online“-Portal „verliehen“. Gedacht wird tatsächlich in Exemplaren: „entliehen“ wird eine mit DRM-Schutz und Zeitvektor versehene Datei (bislang vor allem im PDF-Format, in Zukunft vermehrt auch im EPUB-Format) für eine bestimmte Dauer. Dabei kaufen die Bibliotheken eine bestimmte Anzahl an „Exemplaren“, das heißt wenn eine Datei „verliehen“ ist, kann der Titel vorgemerkt werden. Ein Simultanzugriff, wie er häufig in Wissenschaftlichen Bibliotheken über VPN realisiert wird, ist nicht möglich. Erst nach Ablauf der Leihfrist wird der Zugriff wieder freigegeben. Dass insbesondere „Berufstätige im mittleren Alter“ ein solches Angebot nutzen,[69] verwundert nicht. „Digital natives“ dürfte ein derartiges Modell eher absurd erscheinen.

Mit einem anderen Modell als der Ausleihimitation experimentieren beispielsweise der Verbund Öffentlicher Bibliotheken Berlin (VÖBB) und der Berlin Verlag seit September 2009. Der Berlin Verlag ist auf einer Online-Plattform mit wichtigen Titeln aus seinen Verlagsprogrammen vertreten, die Titel sind im OPAC des Verbunds nachgewiesen und nach Anmeldung für die Nutzerinnen und Nutzer des VÖBB zugänglich. Dabei werden keine Dateien heruntergeladen, sondern die Texte werden im Webbrowser gelesen; ein Simultanzugriff ist daher möglich. Auch im Bereich der Öffentlichen Bibliotheken wären in Zukunft konsortiale Erwerbungsmodelle vorstellbar, die Lösungen für digitale Angebote jenseits des Monopolisten DiViBib

[68] http://www.divibib.com/27.0.html

[69] Das Ergebnis einer Nutzungsstudie aus dem Jahr 2009 lautet: „Zusammengefasst lässt sich daraus folgern, dass die Hauptnutzer berufstätig, im Alter zwischen 36 und 55 sind, einen höheren Bildungsabschluss haben und es keinen geschlechtsspezifischen Unterschied gibt“ (Gudrun Kulzer: „Gut ausgebildete Berufstätige im mittleren Alter liegen vorne. Studie ermittelt Hauptnutzer virtueller Angebote / Vor allem E-Books gefragt“. In: BuB 61 (2009), S. 624-627, Zitat S. 626).

erproben.[70]

In eine ähnliche Richtung, wenn auch mit einer zusätzlichen Speicherfunktionalität zum Offline-Lesen, scheint Google mit seiner Verkaufsplattform Google eBookstore, die zunächst „Google Editions“ hieß, zu gehen, vielleicht auch um die unüberschaubare Vielfalt an – teilweise proprietären – Formaten, die auf dem amerikanischen E-Book-Markt besteht, zu umgehen. Die Titel werden im Webbrowser zugänglich gemacht, zusätzlich wird der Titel im Cache auch offline einsehbar – die Verlage entscheiden, ob sie dabei DRM verwenden möchten oder nicht.[71] Mit dieser Entwicklung bewegt sich Google weiter in Richtung Cloud Computing und Web App. Gekauft wird also nicht mehr die Datei oder das digitale Dokument, sondern lediglich der Zugang. Voraussetzung sind natürlich internetfähige Endgeräte – E-Book-Reader ohne Internetzugang werden jedoch voraussichtlich bald der Vergangenheit angehören, da sie absolut inkompatibel mit dem Charakter der Vernetztheit unter digitalen Bedingungen sind.

Digitale Texte bieten auch für nicht-wissenschaftliche Nutzung einige Vorteile gegenüber dem physischen Buch: der Erwerb beziehungsweise der Zugriff ist jederzeit und auf unkomplizierte Weise möglich, es existieren Funktionalitäten zur Volltextsuche, elektronische Annotationen durch die Leserinnen und Leser können im Web-2.0-Kontext miteinander geteilt werden und mittels Hyperlinks und Integration von Text, Bild und bewegtem Bild eröffnen sich neue Potenziale für experimentelle Literatur. Bislang werden die Möglichkeiten jedoch noch nicht konsequent genutzt beziehungsweise in Emulationsszenarien verhindert. Auch ist fraglich, ob sich Endgeräte wie E-Book-Reader oder Smartphones zum Lesen von Belletristik durchsetzen können. So bleiben Projekte aus dem Bibliotheksbereich zunächst Angebote, die wahrscheinlich vor allem zum Anlesen der Romane dienen. Es bleibt abzuwarten, ob sich in Zukunft auch für belletristische digitale Literatur neue Aggregationsmöglichkeiten und semantische Zugangswege, wie sie in der sechsten These dieser Arbeit skizziert wurden, entwickeln – erste Ansätze oder zumindest

[70] Erste Schritte im Bereich des eLerning in diese Richtung haben die Hamburger Öffentlichen Bücherhallen mit dem Anbieter bit media gemacht – andere Öffentliche Bibliotheken könnten sich anschließen. Einfacher zu koordinieren wären Konsortien natürlich, gäbe es einen nationalen Verbund (Öffentlicher Bibliotheken).

[71] Bislang hat Google noch keine detaillierten Informationen über seine Pläne veröffentlicht. Für einen ersten Überblick vgl. http://books.google.com/ebooks. Ein Angebot zum online-Lesen erworbener Bücher bietet auch heute schon Ciando: http://www.ciando.com/service/online-lesen/

Potenziale zeigen kleine Projekte auf dem deutschen Markt wie das Portal bilandia[72] oder das auf semantischen Technologien basierende Thoughtlab von Europeana[73].

[72] bilandia (http://www.bilandia.de/) bietet eine iPhone-App an, die Suchfunktionen nach Büchern (zum Beispiel per Landkarte, Zeitreise oder nach Farben) mit Web-2.0-Funktionalitäten und Kaufmöglichkeiten verknüpft.

[73] http://www.europeana.eu/portal/thought-lab.html

5 Fazit

In der Übergangsphase vom Gutenberg- zum digitalen Zeitalter verliert der Dokumentbegriff zunächst an Konturschärfe, denn das traditionelle Dokument lässt sich als beschrifteter Träger definieren, der als physisches Objekt klare Grenzen besitzt (vgl. **These 1**). Diese eindeutige Abgrenzung des Dokuments geht in der digitalen Umgebung im Bitstream, der eine Unterscheidung von Signal und Medium nicht mehr zulässt, verloren. Die Übergangsphase vom analogen zum digitalen Paradigma, in der wir uns derzeit noch befinden, ist von Imitationsversuchen geprägt: Das traditionelle Dokument wird durch Emulationsmodi – wie beispielsweise PDF – ins Digitale übersetzt, nicht zuletzt weil die Befürchtung besteht, mit der Erosion des Dokumentbegriffs auch der mit dem Dokument verbundenen Funktionen innerhalb der sozialen Kommunikation verlustig zu gehen (vgl. **These 2**). Allerdings ist es schlicht unmöglich, den traditionellen Dokumentbegriff in die digitale Welt zu übertragen. Dieser Umstand zeigt sich am hier diskutierten Untersuchungsgegenstand – digitaler Belletristik – besonders deutlich, ist damit doch ein Bereich gewählt, der sich – beispielsweise im Gegensatz zu naturwissenschaftlichen Publikationen – besonders stark einem Neudenken im Digitalen widersetzt. Von der literaturwissenschaftlichen Forschung wird belletristische Literatur in digitaler Form noch weitestgehend ignoriert; dementsprechend spielt sie im E-Book-Angebot von Wissenschaftlichen Bibliotheken bislang auch kaum eine Rolle. Die von der Editionswissenschaft geprägte und inzwischen auch in Deutschland institutionalisierte Computerphilologie hat sich in den letzten Jahrzehnten vor allem mit Retrodigitalisierungsprojekten beschäftigt und besitzt mit den Richtlinien der Text Encoding Initiative (TEI) eine international verwendete und dementsprechend Interoperabilität gewährleistende Auszeichnungssprache für literarische Texte, die insbesondere bei der Erstellung digitaler Editionen Anwendung findet. Jedoch scheint es zumindest im deutschen akademischen Raum noch nicht ausreichend Erfahrungen mit der Anwendung von TEI auf genuin digitale Texte zu geben, um einschätzen zu können, ob sich diese komplexen Richtlinien auch für derartige Texte eignen. Es ist jedoch anzunehmen, dass dies nicht der Fall ist, da die TEI-Richtlinien eben zur Enkodierung von Druckvorlagen geschaffen wurden und insofern einem Emulationsdenken verhaftet

bleiben. Für Belletristik als allgemeines Kulturgut und also für Verlage und Dienstleister auf dem Buchmarkt spielen die TEI-Richtlinien keine Rolle. Zwar sind seit einigen Jahren vermehrt Bemühungen auf Seiten der Publikumsverlage festzustellen, digitale Ausgaben ihrer Publikationen anzubieten. Allerdings scheint hier die Entwicklung bislang in die Richtung zu gehen, mit von Dienstleistern geschaffenen DTDs E-Books gemäß dem XML-basierten Standard EPUB anzubieten, die eine Lektüre auf verschiedenen Endgeräten ermöglichen. Allerdings widersetzen sich die mit Kopierschutz versehenen belletristischen E-Books dem digitalen Paradigma und einem seiner Charakteristika – der Vernetzung. In Öffentlichen Bibliotheken, in denen belletristische E-Books bislang fast ausschließlich über die von DiViBib angebotene „Onleihe“ zugänglich gemacht werden, zeigt sich das – zumal noch stark der Kultur des traditionellen Urheberrechts verhaftete – Emulationsdenken besonders deutlich: An physischen Medien orientierte Ausleihmodelle konterkarieren geradezu die digitale Umgebung.

Dabei ist es nicht nur sein Potenzial zur Vernetzung, worin sich das digitale Dokument vom analogen unterscheidet, sondern vor allem auch das Charakteristikum der Diskretheit, das die Voraussetzung jeglicher Art von Datenprozessierung und damit also der Maschinenlesbarkeit von Text darstellt (vgl. **These 3**). Die Möglichkeit zur Prozessierung basiert auf der Trennung von Form (Stylesheet) und Inhalt (XML), was im Vergleich zum traditionellen, durch die Einheit von Form und Inhalt charakterisierten Dokument etwas ganz und gar Neuartiges darstellt. Allerdings gerät dadurch der informatische Textbegriff mit einem hermeneutisch fundierten Text- und Dokumentbegriff in Konflikt, denn ein Dokument ist stets dazu intendiert, von einem menschlichen Interpreten rezipiert zu werden. Die durch die binäre Kodierung eingezogene Signifikatenebene macht die hermeneutische Dimension jedoch keineswegs obsolet, stellt doch insbesondere die für das wissenschaftliche Arbeiten mit Primärtexten bedeutsame Strukturierung der Daten bereits eine hermeneutische Tätigkeit dar. Gleichzeitig erlaubt die Binärkodierung, große digitale Korpora mittels – wiederum hermeneutischer Tätigkeit entstammenden – Algorithmen durchdringen zu lassen.[74] Damit ist ein Schritt vom

[74] In diesem Kontext ist interessant, dass mit Googles Digitalisierungsbemühungen das bislang wahrscheinlich größte digitale Korpus entsteht und Google derzeit ein Forschungsprogramm ins Leben ruft, wie mit Methoden der 'digital humanities' mit diesem Korpus gearbeitet werden kann (vgl. Parry). Mit dem Books Ngram Viewer liegt ein erstes Instrument (http://ngrams.googlelabs.com/info) der sogenannten „Culturomics“ vor (vgl. http://www.culturomics.org/). Vgl. hierzu Michel et al. 2010.

Lesen zum Rechnen der Literatur vollzogen, der schon anzudeuten vermag, warum sich die hermeneutisch fundierten Wissenschaften bislang noch mit digitalen Arbeitsmethoden schwer tun. Auch in Bezug auf die bislang eher zurückhaltende Auszeichnung digitaler Texte bei der Herstellung von E-Books durch belletristische Verlage ist eine solche hermeneutische Tätigkeit zu konstatieren, wobei abzuwarten bleibt, wie sich die Gestaltungskonventionen unter digitalen Bedingungen weiter entwickeln werden.

Die Erosion der klar definierten Dokumentgrenzen unter digitalen Bedingungen – durch Algorithmen, Hyperlinks und Ähnliches – und die Trennung von Form und Inhalt führt dazu, dass zunächst dem Inhalts- beziehungsweise dem Textbegriff gegenüber dem Dokumentbegriff eine größere Bedeutung zukommt (vgl. **These 4**). Diese Fokussierung auf den Inhalt zeigt sich zudem auch an neueren Entwicklungen in Richtung Semantic Web beziehungsweise Linked Data, die darauf abzielen, das „web of documents“ langfristig in ein „web of data“ zu überführen. Zu problematisieren ist dabei allerdings – etwa im Kontext des „ontology mapping“ – das diesen Bemühungen zugrunde liegende Sprachverständnis. Für die wissenschaftliche Nutzung bieten semantische Technologien – wie am Beispiel der digitalen Bibliotheken des Discovery-Projekts gezeigt wurde – jedoch ein großes Potenzial. Mit dem „web of data“ gehen auch neue Publikationsformen einher, insbesondere ist – gerade auch vor der Hintergrund der Schaffung von Virtuellen Forschungsumgebungen für Geisteswissenschaften – zu hoffen, dass analog zur Bereitstellung von Primärdaten in den Naturwissenschaften auch die Veröffentlichung von getaggten Primärtexten für die Philologien Verbreitung findet, denn schließlich ist die philologische Sekundärliteratur meist erst auf Grundlage der Primärtexte angemessen rezipierbar. Wissenschaftlichen Bibliotheken kommt mit dem Aufschwung der 'digital humanities' in völlig neuartiger Qualität die Funktion einer Schulungs- und Serviceeinrichtung für Wissenschaftlerinnen und Wissenschaftler zu. Außerdem werden sie zu wichtigen Partnern der Wissenschaft im Bereich des Daten- und Informationsmanagements. Aber auch Öffentliche Bibliotheken können semantische Technologien zur Verbesserung komplexer Suchen und zur Präsentation digitaler Texte nutzen, wobei bislang – analog zum Buchmarkt – noch kaum Entwicklungen in diese Richtung zu bemerken sind.

Um die neueren digitalen Technologien auch von geisteswissenschaftlicher Seite

mitgestalten zu können, ist es wichtig, mit der Vorstellung von neutraler Technik zu brechen. Technischen Formaten und Standards liegen stets Text- und Dokumentmodelle zugrunde, die Auswirkungen auf zukünftige Arbeits- und Lektüremöglichkeiten haben. Während durch die reine Binärkodierung von Texten ein simplifizierender Textbegriff befördert wird, kann die Strukturierung der Daten durch Auszeichnungssprachen einen hermeneutisch fundierten Textbegriff wiederherstellen (vgl. **These 5**). So führt beispielsweise die Konzeption der TEI-Richtlinien auf der Basis eines Modells von Text als 'Ordered Hierarchy of Content Objects' (OHCO) nicht nur zu theoretischen Debatten, sondern auch zu Schwierigkeiten in der Praxis der Textenkodierung (etwa dem Problem der 'overlapping hierarchies'). Nichtsdestotrotz ist festzustellen, dass die Möglichkeit zur Textauszeichnung im Digitalen dem Text einen entscheidenden Mehrwert hinzufügt: In XML können unterschiedliche Interpretationen eines Textes in den Primärtext selbst einfließen beziehungsweise mit ihm gemeinsam präsentiert werden, womit sich neue Möglichkeiten zur Dokumentation der Rezeptionsgeschichte eines Textes ergeben.

Nach der Diagnose einer Erosion des Dokumentbegriffs und mit der Analyse der neuen technischen Möglichkeiten stellt sich zugleich die Frage nach einer Rekonstruktion des Dokumentbegriffs unter digitalen Bedingungen. Deutlich wird, dass sich die mit dem Dokumentbegriff verbundenen wissenschaftlichen und kulturellen Funktionalitäten nicht in Emulationsszenarien sicherstellen lassen, sondern nur unter konsequenter Beachtung des digitalen Paradigmas. Dabei vermag unter digitalen Bedingungen manches Desiderat eingelöst zu werden, das gedruckte Texte nicht erfüllen konnten, auch wenn avantgardistische Schreibweisen die nunmehr offen stehenden Möglichkeiten teilweise bereits vorweggenommen haben (vgl. **These 6**). Im wissenschaftlichen Bereich zeigen digitale Editionen und Werkzeuge bereits mögliche zukünftige Entwicklungslinien auf: Methoden des 'distant reading', der Intersemiose, der Nachnutzung von Texten und des kollaborativen Arbeitens, der Vernetzung und Intertextualität entstehen. Einen Ansatz, auf welche Weise sich in einer derartig offenen und vernetzten Umgebung trotzdem Entitäten herausbilden können, die dazu in der Lage sind, die stabilisierenden und autorisierenden Funktionen des Dokuments aufzuheben, zeigt beispielsweise die Open Archives Initiative Object Reuse and Exchange (OAI-ORE) auf. Basierend auf RDF werden Ressource-Aggregationen gebildet und in einer

'resource map' dokumentiert. Sowohl für den wissenschaftlichen Umgang mit literarischen Texten als auch für den Zugang zu Belletristik im Kulturbereich hat ein solches flexibles und zugleich stabilisierendes Modell Vorteile: Verschiedene Interpretationsschichten können einem Primärtext hinzugefügt, einzelne Textteile nachgenutzt werden, neue Zusammenstellungen von belletristischen Texten sind einfach und vergleichsweise umstandslos zu realisieren, vor allem – dank RDF – über semantische Zugangsmöglichkeiten und Vernetzungen.

Die Rolle von Bibliotheken, soviel ist schon heute zu spüren, verändert sich unter digitalen Bedingungen fundamental. Bibliotheken bemühen sich, neue Zugangsmöglichkeiten zu Inhalten und Informationen zu bieten. In diesem Kontext ist ein Verständnis der neueren Technologien und Entwicklungen und eine Kooperation mit Entwicklern essentiell. Neben dem Hosten der Daten und ihrer Langzeitarchivierung, die Bibliotheken mit Archivfunktion übernehmen müssen, besteht die vornehmliche Funktion von Bibliotheken darin, die Suche nach Informationen und den Zugriff auf Inhalte einfach zu gestalten. Wissenschaftlichen und Öffentlichen Bibliotheken kommen dabei jedoch auch vermehrt Aufgaben in der Vermittlung von Techniken wissenschaftlichen Arbeitens unter digitalen Bedingungen und des elektronischen Publizierens beziehungsweise einer sinnvollen und legalen Nutzung des WWW zu. Eine interessante neue Herausforderung wird dabei für Bibliotheken in der Moderation zwischen Entwicklern und Informatikern auf der einen Seite und Wissenschaftlern beziehungsweise ihren Nutzerinnen und Nutzern und kulturellen Institutionen auf der anderen Seite bestehen.

Zwar wird das WWW als neuer Vertriebsweg von Belletristik bislang – zumindest auf Rezeptionsseite – lediglich von einer relativ kleinen Gruppe und zudem ohne die traditionelle Selektionsfunktion des Verlagslektorats genutzt, sodass reinen Netzpublikationen belletristischer Literatur bislang für Bibliotheken eine eher untergeordnete Bedeutung zukommt. Dieser Befund erklärt im Übrigen, weshalb genuine Netzliteratur im Rahmen dieser Arbeit keine Berücksichtigung findet. Dennoch ist zu erwarten, dass sich in Zukunft auch auf diesem Feld des Literaturbetriebs Publikationsformen ausbilden, die – analog zu Peer-Review-Verfahren wissenschaftlicher E-Journals – einer qualitativen Auswahl unterzogen werden. Auch wenn es wünschenswert ist, die Archivierung von Netzpublikationen

zentral zu organisieren,[75] werden sich jedoch alle Bibliotheken mit der Erschließung digitaler Literatur beziehungsweise ihrer Präsentation in lokalen Sammlungszusammenhängen auseinandersetzen müssen. Nicht jede dieser Veröffentlichungen wird wie jene Texte aus Rainald Goetz' Blog „Abfall für alle“, die nach einem Jahr im Suhrkamp Verlag erschienen sind, als gedrucktes Buch fixiert werden. Auf derartige Entwicklungen zu reagieren, die Archivierung dieser Inhalte zu gewährleisten, eine Auswahl für spezielle Rezipientengruppen zu treffen und neue Zugangswege zu entwickeln, wird in Zukunft dementsprechend eine weitere Herausforderung für Bibliotheken darstellen.

[75] Vgl. http://www.d-nb.de/netzpub/index.htm

Literaturverzeichnis

“2009 Librarian eBook Survey.” HighWire Press, 2010. http://highwire.stanford.edu/PR/HighWireEBookSurvey2010.pdf.

Barbera, Michele, Michele Nucci, und Daniel Hahn. “A Semantic Web Powered Distributed Digital Library System,” 2008. http://elpub.scix.net/cgi-bin/works/Show?_id=130_elpub2008&sort=DEFAULT&search=barbera&hits=3.

Benjamin, Walter. “Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit [Erste Fassung].” In *Walter Benjamin: Gesammelte Schriften*, I/2:431-469. Frankfurt am Main: Suhrkamp Verlag, 1991.

Berners-Lee, Tim, und James Hendler. “Publishing on the semantic web.” *Nature* 410, no. 6832 (April 26, 2001): 1023-1024.

———. “Publishing on the semantic web.” *Nature* 410, no. 6832 (April 26, 2001): 1023-1024.

Berners-Lee, Tim, James Hendler, und Ora Lassila. “The semantic web.” *Scientific American* 284, no. 5 (2001): 34-43.

Biller, Maxim. *Wenn ich einmal reich und tot bin. Erzählungen*. Köln: Kiepenheuer & Witsch, 1990.

Bizer, Christian, Tom Heath, und Tim Berners-Lee. “Linked data–the story so far.” *International Journal on Semantic Web and Information Systems* 5, no. 3 (2009): 1-22.

Blumauer, Andreas, und Tassilo Pellegrini. “Semantic Web und semantische Technologien: Zentrale Begriffe und Unterscheidungen.” In *Semantic Web*, 9-25. Berlin, Heidelberg: Springer, 2006. http://dx.doi.org/10.1007/3-540-29325-6_2.

Booth, David, Hugo Haas, Francis McCabe, Eric Newcomer, Michael Champion, Chris Ferris, und David Orchard. “Web Services Architecture,” 2004. http://www.w3.org/TR/ws-arch/.

Briet, Suzanne. “Bibliothécaires et documentalistes.” *Revue de la Documentation* 21 (1954): 41-45.

———. *Qu'est-ce que la documentation?* Paris: EDIT, 1951.

Buckland, Michael. “What is a "digital document"?” (1998). http://people.ischool.berkeley.edu/~buckland/digdoc.html.

Burch, Thomas, Claudine Moulin, und Andrea Rapp. “Informatik in den Geisteswissenschaften.” *it - Information Technology* 51, no. 4 (7, 2009): 181-182.

Burnard, Lou, und Syd Bauman, Hrsg. *TEI P5: Guidelines for electronic text encoding and interchange*, 2009. http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf.

Burnard, Lou, und C. M. Sperberg-McQueen. “TEI Lite: Encoding for Interchange: an introduction to the TEI — Revised for TEI P5 release,” 2006. http://www.tei-

c.org/release/doc/tei-p5-exemplars/html/teilite.doc.html.

Buzzetti, Dino. “Digital Representation and the Text Model.” *New Literary History* 33, no. 1 (Winter 2002): 61-88.

Capurro, Rafael. “Hermeneutik revisited.” *Semantic Web. Wege zur vernetzten Wissensgesellschaft. Springer Berlin, Heidelberg* (2006): 527-533.

Carusi, Annamaria, und Torsten Reimer. “Virtual Research Environment. Collaborative Landscape Study. A JISC funded Project,” 2010. http://www.jisc.ac.uk/publications/reports/2010/vrelandscapestudy.aspx#downloads.

Chervel, Thierry. “Fantasie über die Zukunft des Schreibens,” März 14, 2010. http://www.perlentaucher.de/blog/134_fantasie_ueber_die_zukunft_des_schreibens.

Connolly, Dan. “Gleaning Resource Descriptions from Dialects of Languages (GRDDL),” 2007. http://www.w3.org/TR/grddl/.

Cramer, Florian. “Digital Code and Literary Text,” 2001. http://www.dvara.net/Hk/digital_code_and_literary_text.pdf.

Crane, Gregory. “What Do You Do with a Million Books?.” *D-Lib Magazine* 12, no. 3 (3, 2006). http://www.dlib.org/dlib/march06/crane/03crane.html.

Day, Ronald E. “"A Necessity of Our Time": Suzanne Briet's What is Documentation?.” *Document (re) turn. Contributions from a research field in transition* (2007): 311-326.

DeRose, S. J, D. G Durand, E. Mylonas, und A. H Renear. “What is text, really?.” *Journal of Computing in Higher Education* 1, no. 2 (1990): 3-26.

Deutsche Forschungsgemeinschaft. “60.06 - Merkblatt. Service-Projekte zu Informationsmanagement und Informationsinfrastruktur in Sonderforschungsbereichen INF” (2009). http://www.dfg.de/download/programme/sonderforschungsbereiche/allgemeine_hinweise/60_06/60_06.pdf.

———. “Wissenschaftliche Literaturversorgungs- und Informationssysteme: Schwerpunkte der Förderung bis 2015. DFG-Positionspapier” (2006). http://www.dfg.de/download/pdf/foerderung/programme/lis/positionspapier.pdf.

Dubin, D., A. Renear, C. M. Sperberg-McQueen, und C. Huitfeldt. “A logic programming environment for document semantics and inference.” *Literary and Linguistic Computing* 18, no. 1 (2003): 39-47.

Dworschak, Manfred. “Babylonischer Bau. Die Deutsche Digitale Bibliothek will Millionen Bücher, Filme, Bilder und Tonaufnahmen im Internet zugänglich machen. Über 30000 Bibliotheken, Museen und Archive sollen ihr digitalisiertes Kulturgut beisteuern. Kann ein derart ehrgeiziger Plan überhaupt gelingen?.” *Der Spiegel*, no. 6 (2010): 142-144.

Eco, Umberto. *Zeichen: Einführung in einen Begriff und seine Geschichte*. Übersetzt von Günter Memmert. 1. Aufl. Frankfurt am Main: Suhrkamp, 1977.

Foucault, Michel. *Dies ist keine Pfeife*. München, Wien: Hanser, 1997.

“Für Publikationsfreiheit und Wahrung der Urheberrechte [Heidelberger Appell],”

2009. http://www.textkritik.de/urheberrecht/index.htm.

Genette, Gérard. *Palimpseste: die Literatur auf zweiter Stufe*. 3. Aufl. Frankfurt am Main: Suhrkamp, 2001.

Giesecke, Michael. *Sinnenwandel, Sprachwandel, Kulturwandel: Studien zur Vorgeschichte der Informationsgesellschaft*. 1. Aufl. Frankfurt am Main: Suhrkamp, 1992.

Golumbia, David. *The Cultural Logic of Computation*. Cambridge, Mass.: Harvard University Press, 2009.

Gradmann, Stefan. “Building Blocks of the future Scholarly Web: Beyond and far beyond. The APE Lecture 2010,” Berlin, 2010. http://river-valley.tv/building-blocks-of-the-future-scholarly-web-beyond-and-far-beyond/.

———. “Open Access-einmal anders. Zum wissenschaftlichen Publizieren in den Geisteswissenschaften.” *Zeitschrift für Bibliothekswesen und Bibliographie* 54, no. 4/5 (2007): 170.

———. “rdfs:frbr - Towards an Implementation Model for Library Catalogs Using Semantic Web Technology.” *Cataloging & Classification Quarterly* 39, no. 3/4 (2005): 63-75.

———. *Signal, Information, Zeichen: zu den Bedingungen des Verstehens in semantischen Netzen; Antrittsvorlesung 28. Oktober 2008*. Öffentliche Vorlesungen / Humboldt-Universität zu Berlin; 157. Berlin: Humboldt-Universität, 2009. http://nbn-resolving.de/urn:nbn:de:kobv:11-10096931.

———. “Vom Verfertigen der Gedanken im digitalen Diskurs: Versuch einer wechselseitigen Bestimmung hermeneutisch und empirizistischer Positionen.” *Historical Social Research* 29, no. 1; ISSU 107 (2004): 56-63. http://hsr-trans.zhsf.uni-koeln.de/hsrretro/docs/artikel/hsr/hsr2004_598.pdf.

Gradmann, Stefan, und Jan Christoph Meister. “Digital document and interpretation: re-thinking “text” and scholarship in electronic settings.” *Poiesis & Praxis: International Journal of Technology Assessment and Ethics of Science* 5, no. 2 (2008): 139-153.

Gruber, Tom. “Ontology.” *Encyclopedia of Database Systems. Ling Liu and M. Tamer Özsu (Eds.), Springer-Verlag* (2009).

van Harmelen, Frank. “Semantic Web Research anno 2006: main streams, popular fallacies, current status and future challenges.” *Cooperative Information Agents X. 10th International Workshop, CIA 2006 Edinburgh, UK, September 11-13, 2006 Proceedings. Springer Berlin/Heidelberg* (2006): 1-7.

Horn, Dara. *Vor allen Nächten: Roman. Aus dem Amerikan. von Christiane Buchner und Martina Tichy*. Berlin: Berlin-Verlag, 2009.

IFLA Study Group on the Functional Requirements for Bibliographic Records, Hrsg. *Functional Requirements for Bibliographic Records*, 1997. http://www.ifla.org/VII/s13/frbr/.

Jacobs, Ian, und Norman Walsh. “Architecture of the World Wide Web,” 2004. http://www.w3.org/TR/webarch/.

Jannidis, Fotis. “Computerphilologie.” In *Handbuch der Literaturwissenschaft. Drei Bände. Bd II: Methoden und Theorien*, Stuttgart & Weimar: JB Metzler, 2007:

27-40.

———. *Computerphilologie*. Stuttgart: Metzler, 2009.

———. “TEI in a crystal ball.” *Lit Linguist Computing* 24, no. 3 (September 1, 2009): 253-265.

Kelly, Kevin. “Scan This Book!.” *The New York Times*, Mai 14, 2006. http://www.nytimes.com/2006/05/14/magazine/14publishing.html?_r=1.

Kittler, Friedrich. *Aufschreibesysteme : 1800 - 1900.* 3. Aufl. München: Wilhelm Fink, 1995.

———. “Von der Letter zum Bit.” In Horst Wenzel, Friedrich Kittler, Manfred Schneider (Hg.): *Gutenberg und die neue Welt,* München: W. Fink 1994: 105-117.

Klyne, Graham, und Jeremy J. Carroll. “Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation,” 2004. http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/.

Knobloch, Clemens. “Text/Textualität.” In Karlheinz Barck u. a. (Hg.): *Ästhetische Grundbegriffe. Historisches Wörterbuch in sieben Bänden*, Bd. 6, Stuttgart [u.a.]: Metzler, 2005: 23-48.

Kristeva, Julia. “Bachtin, das Wort, der Dialog und der Roman.” *Literaturwissenschaft und Linguistik. Ergebnisse und Perspektiven* 3 (1972): 345-375.

Lagoze, Carl, und Herbert Van de Sompel. “ORE User Guide - Primer,” 2008. http://www.openarchives.org/ore/1.0/primer.

Lagoze, Carl, Herbert Van de Sompel, Michael Nelson, Simeon Warner, Robert Sanderson, und Pete Johnston. “A Web-Based Resource Model for eScience: Object Reuse & Exchange.” *0811.0573* (November 4, 2008). http://arxiv.org/abs/0811.0573.

Lauer, Gerhard. “Literatur rechnen: Lektüre im Computerzeitalter.” *FAZ.NET*, August 26, 2009. http://www.faz.net/-00mu6w.

Lopez, P. “GROBID: Combining Automatic Bibliographic Data Recognition and Term Extraction for Scholarship Publications.” In *Proceedings of ECDL*, 2009.

Lund, Niels Windfeld. “Documentation in a complementary perspective.” In *Aware and Responsible: Papers of the Nordic-International Colloquium on Social and Cultural Awareness and Responsibility in Library, Information and Documentation Studies (SCARLID)*, 93-102, 2004.

Lund, Niels Windfeld, und Michael Buckland. “Document, documentation, and the Document Academy: introduction.” *Archival Science* 8, no. 3 (2008): 161-164.

Mahoney, Michael S. “The histories of computing(s).” *Interdisciplinary Science Reviews* 30, no. 2 (2005): 119-135.

McCarty, Willard. *Humanities computing*. 1. Aufl. Houndmills [u.a.]: Palgrave Macmillan, 2005.

McGann, Jerome. “Rethinking Textuality.” In *Radiant Textuality: Literature after the World Wide Web*. New York, Hampshire: Palgrave Macmillan 2001: 137-166. http://www2.iath.virginia.edu/jjm2f/old/jj2000aweb.html.

Meister, Jan Christoph. *Computing action: a narratological approach*. Walter de Gruyter, 2003.

———. “Projekt Computerphilologie” (2003). http://edocs.ub.uni-frankfurt.de/volltexte/2008/11645/pdf/MEISTER_Projekt.pdf.

Michel, J. B, Y. K Shen, A. P Aiden, A. Veres, M. K Gray, B. Team, J. P Pickett, D. Hoiberg, D. Clancy, und P. Norvig. “Quantitative Analysis of Culture Using Millions of Digitized Books.” *Sciencexpress* (Dezember 16, 2010).

Neffe, Jürgen. “Es war einmal. Die Ära des gedruckten Buches geht zu Ende. Kein Grund zur Trauer.” *DIE ZEIT*, April 24, 2009. http://www.zeit.de/2009/18/L-Buch?page=all.

Nelson, Theodor Holm. *Geeks Bearing Gifts. How the computer world got this way*. o. O.: Mindful Press, 2009.

———. “I DON'T BUY IN,” o. J . http://ted.hyperland.com/buyin.txt.

———. “Toward A Deep Electronic Literature: The Generalization of Documents and Media,” 2007. http://xanadu.com/XanaduSpace/xuGzn.htm.

———. “Xanalogical structure, needed now more than ever: parallel documents, deep links to content, deep versioning, and deep re-use.” *ACM Computing Surveys (CSUR)* 31, no. 4es (1999). http://www.signallake.com/innovation/XanalogicalStructure1299.pdf.

Nelson, Theodor Holm, und Robert Adamson Smith. “Back to the Future: Hypertext the Way It Used To Be.” http://xanadu.com/XanaduSpace/btf.htm.

Neuroth, Heike, Fotis Jannidis, Andrea Rapp, und Felix Lohmeier. “Virtuelle Forschungsumgebungen für e-Humanities. Maßnahmen zur optimalen Unterstützung von Forschungsprozessen in den Geisteswissenschaften.” *BIBLIOTHEK Forschung und Praxis* 33, no. 2 (9, 2009): 161-169.

Open Archives Initiative Object Reuse and Exchange. “ORE Specifications and User Guides,” 2008. http://www.openarchives.org/ore/1.0/toc.html.

Ørom, Anders. “The concept of information versus the concept of document.” A d*ocument (re)turn: contributions from a research field in transition* (2007): 53-72.

Otlet, Paul. *Monde: essai d’universalisme: connaissance du monde, sentiment du monde, action organisée et plan du monde*. Brüssel: Editiones Mundaneum, 1935. http://www.laetusinpraesens.org/uia/docs/otlet_contents.php.

———. *Traité de documentation : le livre sur le livre : théorie et pratique.* Bruxelles: Centre de lecture publique de la Communauté française de Belgique, 1989.

Parry, Marc. “Google Starts Grant Program for Studies of Its Digitized Books.” *The Chronicle of Higher Education*, März 31, 2010. http://chronicle.com/article/Google-Starts-Grant-Program/64891/.

Pédauque, Roger T. *La redocumentarisation du monde*. Caen: C&F éditions, 2007.

———. *Le document à la lumière du numérique*. Caen: C&F éditions, 2006.

Posth, Sebastian. “E-Books, Reader und ihre Formate. Anmerkungen zur Zukunft elektronischer Publikationen.” *BuB* 61, no. 09 (2009): 617-619.

Puschmann, Cornelius. “Vom Object Web zum Discourse Web. Metaphern der digitalen Kommunikation im Wandel und ihre Auswirkungen auf die Wissenschaft.” *LIBREAS. Library Ideas* Jg. 5, H. 2 (2009). http://www.libreas.eu/ausgabe15/texte/003.htm.

Renear, A., D. Dubin, und C. M. Sperberg-McQueen. “Towards a semantics for XML markup.” In *Proceedings of the 2002 ACM symposium on Document engineering*, 126, 2002.

Renear, A., D. Dubin, C. M. Sperberg-McQueen, und C. Huitfeldt. “XML semantics and digital libraries.” In *Proceedings of the 3rd ACM/IEEE-CS joint conference on Digital libraries*, 303-305, 2003.

Renear, A. H, E. Mylonas, und D. Durand. “Refining our notion of what text really is: The problem of overlapping hierarchies.” (1993). http://www.ideals.illinois.edu/bitstream/handle/2142/9407/RefiningOurNotion.pdf?sequence=2.

Renear, Allen. “Out of praxis: Three (meta)theories of textuality.” In Kathryn Sutherland (Hg): *Electronic Text. Investigations in Method and Theory,* Oxford: Claderon Press, 1997: 107-126.

———. “Text encoding.” *A companion to digital humanities* (2004): 218-237.

Renear, Allen, Jerome McGann, und Susan Hockey. “Abstract for ACHALLC99 debate: What is Text?,” 1999. http://www.humanities.ualberta.ca/Susan_Hockey/achallc99.htm.

Rockwell, G. “What is text analysis, really?.” *Literary and linguistic computing* 18, no. 2 (2003): 209.

Roesler-Graichen, Michael. *Gutenberg 2.0: die Zukunft des Buches ; eine aktueller Reader zum E-Book*. Herausgegeben von Ronald Schild. Frankfurt am Main: Börsenblatt [u.a.], 2008.

Romary, Laurent. “Questions & Answers for TEI Newcomers,” 2009. http://hal.archives-ouvertes.fr/hal-00348372/en/.

Salaün, Jean-Michel. “Web, texte, conversation et redocumentarisation.” Actes de conférence / Conference Proceedings. https://papyrus.bib.umontreal.ca/jspui/handle/1866/2226.

Salaün, Jean-Michel, und J. Charlet. “Introduction: Comprendre et maîtriser la redocumentarisation du monde” (2007). https://papyrus.bib.umontreal.ca/jspui/bitstream/1866/725/1/Salaun-JM-Redocumentarisation.pdf?language=en.

Schmidt, Arno. *Vorläufiges zu Zettels Traum*. Frankfurt am Main: S. Fischer Verlag, 1977.

Schmidt, Arno, und Arno-Schmidt-Stiftung. *Werkgruppe IV. Das Spätwerk. Julia, oder die Gemälde*. Bd. 4. 1. Aufl. Eine Edition der Arno Schmidt Stiftung. Zürich: Haffman, 1992.

Schreibman, Susan. “Computer-mediated Texts and Textuality: Theory and Practice.” *Computers and the Humanities* 36, no. 3 (2002): 283-293.

———. “The Text Encoding Initiative. An Interchange Format Once Again.” Bd. 10, 2009. http://computerphilologie.tu-darmstadt.de/jg08/schreibman.html.

Shotton, D. “Semantic Publishing: The coming revolution in scientific journal publishing.” *Learned Publishing* 22, no. 2 (2009): 85-94.

Skare, Roswitha, Niels Windfeld Lund, und Andreas Vårheim. *A document (re)turn: contributions from a research field in transition*. Frankfurt am Main [u.a.]: Lang, 2007.

de Sousa, Maria Clara Paixao. “Digital Text: Conceptual and Methodological Frontiers.” *Literatures in the digital era: theory and praxis* (2007): 235-252.

Van de Sompel, Herbert. “Memento: Time Travel for the Web,” 2009. http://www.oclc.org/research/news/2009-12-02.htm.

Voß, Jakob. “Zur Neubestimmung des Dokumentbegriffs im rein Digitalen.” *LIBREAS. Library Ideas* Jg. 5, H2 (15) (2009). http://www.libreas.eu/ausgabe15/texte/002.htm.

W3C. “Semantic Web,” 2010. http://www.w3.org/standards/semanticweb/.

Wehde, Susanne. *Typographische Kultur: eine zeichentheoretische und kulturgeschichtliche Studie zur Typographie und ihrer Entwicklung*. Studien und Texte zur Sozialgeschichte der Literatur ; 69. Tübingen: Niemeyer, 2000.

Weinberger, David. “What's a Document.” *Wired* 4, no. 9 (1996). http://www.wired.com/wired/archive/4.08/document.html.

Witt, Andreas. “Multiple hierarchies: new aspects of an old solution.” 2006. http://opus.kobv.de/ubp/volltexte/2006/865/.

Wolf, Gary. “The Curse of Xanadu,” 1995. http://www.wired.com/wired/archive//3.06/xanadu_pr.html.

Zacklad, Manuel. “Réseaux et communautés d'imaginaire documédiatisée.” *Document (re)turn: contributions from a research field in transition* (2007): 279-297.

Letzter Zugriff auf Hyperlinks: Januar 2011