

# Die Zukunft der Buchstaben in der alphanumerischen Gesellschaft. Text und Dokument unter digitalen Bedingungen<sup>1</sup>



**Sarah Dudek**

Zentral- und Landesbibliothek Berlin  
Fachbereichsleitung Literatur/Sprache  
Blücherplatz 1  
D-10961 Berlin  
Email: dudek@zlb.de

Jahrhunderte lang waren Buch und Text untrennbar miteinander verbunden. Im Bitstream, mit E-Books und digitalen Textkorpora geht diese eindeutige Abgrenzung des Dokuments verloren, und Dokumentkonzepte des Gutenberg-Zeitalters werden unscharf. Es wird dargestellt, wie digitale Umgebungen den traditionellen Dokumentbegriff auflösen und welche Auswirkungen die technologischen Veränderungen auf die Methoden der Geisteswissenschaften und somit auch auf Verlage und insbesondere auf die Rolle von Bibliotheken haben. Der Beitrag schließt mit einem Ausblick, wie ein genuin digital gedachtes Dokument aussehen könnte.

**Schlüsselwörter:** Bibliothek; Dokument; elektronische Publikation

## The Future of Letters in the Alphanumeric Society. Text and Document under Digital Conditions

For centuries books and texts could not be separated from each other. Documents were distinctly delimited entities. In times of bitstream, World Wide Web, and digital text corpora the boundaries of documents as well as the concept of document itself have become vague. The technical and conceptional developments of the last decades lead to changing methods in humanities, publishing houses, and libraries.

**Keywords:** Document; digital library; electronic publishing

## Inhaltsübersicht

0	Einleitung .....	189
1	These zur Erosion des Dokumentbegriffs .....	190

1 Der Titel ist in Anlehnung an Vilém Flussers Aufsatz „Alphanumerische Gesellschaft. Die Zukunft des Buchstabenlesens“ aus dem Jahr 1989 zu verstehen. In: Flusser, Vilém: Die Revolution der Bilder: der Flusser-Reader zu Kommunikation, Medien und Design. Mannheim 1995, S. 38-58.

Der Artikel basiert auf Ergebnissen der Masterarbeit der Verfasserin „Schöne Literatur binär kodiert: Die Veränderung des Text- und Dokumentbegriffs am Beispiel digitaler Belletristik und die neue Rolle von Bibliotheken“. Berlin 2011 (Berliner Handreichungen zur Bibliotheks- und Informationswissenschaft; 290), <http://edoc.hu-berlin.de/series/berliner-handreichungen/2011-290> oder <http://hdl.handle.net/10760/15548>. Die Verfasserin dankt dem Verein zur Förderung der Informationswissenschaft (VFI) für den Anerkennungspreis, mit dem er diese Arbeit 2011 ausgezeichnet hat.

2	These zum Wandel der Kulturtechniken .....	190
3	These zur Diskretheit .....	191
4	These zur Rolle des Textbegriffs .....	192
5	These zu technischen Formaten und Standards ...	194
6	These zur Rekonstruktion des Dokumentbegriffs ...	197

## 0 Einleitung

Dass sich das Wort „Buchstabe“ – wie im Grimmschen Wörterbuch nachzulesen – tatsächlich von den „*büchnen stäbe[n]*“ herleitet,<sup>2</sup> in die die Germanen ihre Schriftzeichen hineinritzten, ist längst in Zweifel gezogen worden. Der Begriff entstand wohl als Bezeichnung für die im Buch verwandten lateinischen Schriftzeichen.<sup>3</sup> Im digitalen Zeitalter, in Zeiten alphanumericer Codierung, führen jedoch beide Herleitungsversuche gleichermaßen ins Nichts; sie sind, wie es Friedrich Nietzsche formuliert, „Metaphern, die abgenutzt und sinnlich kraftlos geworden sind, Münzen, die ihr Bild verloren haben und nun als Metall, nicht mehr als Münzen in Betracht kommen“.<sup>4</sup>

Der Hort der Buchstaben sind die Dokumente. Dokumente besitzen seit Jahrhunderten eine konstitutive Funktion in der gesellschaftlichen Kommunikation. Der Umgang mit ihnen folgt festgeschriebenen oder tradierten Regeln: vom Urheberrecht bis zu Kulturtechniken wie Lesen, Schreiben, Zitieren und Publizieren. Das Dokument spielt in unterschiedlichen Bereichen eine zentrale Rolle: Verlage treten als Produzenten von Dokumenten auf. Für Geisteswissenschaften stellen Dokumente den Haupt-Untersuchungsgegenstand dar. Bibliotheken organisieren Dokumente, machen sie zugänglich und halten sie verfügbar.

Ausgangspunkt für die folgenden Thesen ist die Beobachtung, dass das Dokument selbst und dementsprechend auch das traditionelle Verständnis von Dokument unter digitalen Bedingungen radikalen Veränderungen ausgesetzt sind.<sup>5</sup> Wenn sich das Dokument unter digitalen Bedingungen verändert, bedeutet das auch, dass der Publikationsprozess im Wandel ist – wie auch die Methoden der Wissenschaft und die Arbeit und Aufgaben von Bibliotheken.

2 Vgl. Grimm, Jakob; Grimm, Wilhelm: Das deutsche Wörterbuch. Leipzig, 1854-1971, Bd. 2, Sp. 467 und 480, <http://pom.bbaw.de/dwb/> [Stand: 18.12.2011].

3 Vgl. z.B. Pfeifer, Wolfgang: Etymologisches Wörterbuch des Deutschen. 5. Aufl., München 2000, S. 179f.; Wahrig, Gerhard: Deutsches Wörterbuch : mit einem Lexikon der Deutschen Sprachlehre. 7. Aufl. Gütersloh/München 2001, S. 304.

4 Nietzsche, Friedrich: „Ueber Wahrheit und Lüge im aussermoralischen Sinne“. In: ders.: Sämtliche Werke. Kritische Studienausgabe in 15 Bänden, hrsg. von Giorgio Colli und Mazzino Montinari, Bd. 1, München 1999, S. 873-890, S. 881, <http://www.nietzsche-source.org/texts/eKGWB/WL> [Stand 18.12.2011].

5 Dabei werden Audio-, Film-, 3D- und Bilddokumente zugunsten einer Fokussierung auf Textdokumente außer Acht gelassen.

## 1 These zur Erosion des Dokumentbegriffs

*Das Dokument verliert unter digitalen Bedingungen seine physische Abgrenzung – und Dokumentkonzepte, die sich in der Gutenberg-Ära verfestigt haben, ihre Definitionskraft.*

Als Dokumente können Objekte gelten, die von Menschen zum Zweck der Kommunikation eines Sachverhalts durch Raum und Zeit genutzt und/oder geschaffen werden.<sup>6</sup> Dokumente unterliegen dabei einer permanenten gesellschaftlichen Konsensfindung, die sie mit Wert und Autorität ausstattet. In der von Paul Otlet begründeten Dokumentationswissenschaft hat sich spätestens mit Suzanne Briets 1951 veröffentlichtem Manifest „Qu'est-ce que la documentation?“ ein erweiterter Dokumentbegriff etabliert, der den Fokus von Texten auf Bilder, audiovisuelles Material und Museumsobjekte ausdehnt – bis hin zum Tier als physisches Zeugnis im Kontext eines spezifischen Symbolsystems.<sup>7</sup> Diese Traditionslinie aufgreifend und zusammenfassend definiert Niels W. Lund „documentation“ als Tätigkeit, für die „some kind of instrument is needed“, an der Menschen beteiligt sind und in der eine „meaningful totality“ hergestellt wird.<sup>8</sup>

Um jedoch den Kommunikationsprozess von Gesellschaften zu strukturieren, hat sich über Jahrhunderte hinweg ein Konzept von Dokument tradiert und verfestigt, das eng an das physische Medium gekoppelt ist: an das Papier oder weiter gefasst den Beschreibstoff. Das physische Medium – allen Publikationsformen voran das Buch – ermöglicht den Umgang mit klar abgegrenzten Entitäten. Das traditionelle Dokument ist als Gegenstand zu begreifen und gewährleistet als physisches Objekt seit Jahrhunderten eine große Stabilität. Im Spannungsfeld von Kreation, Distribution und Evaluation von Dokumenten spielen wissenschaftliche Institutionen oder künstlerische Gruppierungen, Verlage und Bibliotheken Hand in Hand und wirken an einem Kanonisierungsprozess mit, der – ganz physisch gedacht – die Masse an Dokumenten über den Publikationsmarkt, gesellschafts-

liche Bewertungssysteme auf zu archivierende Bibliotheksbestände konzentriert.

Bibliotheken sind für diese relativ stabilen Objekte konzipiert worden. Im Übergang vom analogen zum digitalen Signal verliert die physische Form jedoch ihre Definitionskraft für das Dokument: Im Digitalen ist alles ein Bitstream. Mit den Grenzen des physischen Dokuments verschwindet auch die Klarheit des Dokumentbegriffs. Der Bitstream kann nicht mehr – in Analogie zum Papier – als Trägermedium aufgefasst werden, ebenso wenig wie Hardware, Datei oder Bildschirm.<sup>9</sup> Der Begriff des Mediums besitzt in Bezug auf digitale Dokumente grundsätzlich kaum mehr Aussagekraft und sollte daher vermieden werden: Signal und Medium sind im Digitalen untrennbar verbunden.<sup>10</sup>

## 2 These zum Wandel der Kulturtechniken

*Die zahlreichen Imitationen des analogen Mediums im Digitalen zeigen, dass wir uns derzeit in einer Übergangsphase vom analogen zum digitalen Paradigma befinden, in der jedoch die Analogien zunehmend brüchig werden.*

Prozesse des Wandels von Kulturtechniken sind anfänglich generell vom Denken in Analogien geprägt: „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 vervollkommen“.<sup>11</sup> Der Übergang vom analogen zum digitalen Zeitalter stellt dabei keine Ausnahme dar: Das physische Dokument wird imitiert. Dementsprechend sind in der ersten Phase des Wandels vom analogen zum digitalen Paradigma druckanaloge Formate wie PDF zentral geworden. Sie ermöglichen es, den analogen Publikationsprozess im Digitalen nachzubilden, führen jedoch auch zu mancher Paradoxie. So kritisiert Shotton, PDF-Dokumente seien „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“.<sup>12</sup> Nichtsdestotrotz werden

6 Engeren Dokumentkonzepten zufolge können als Dokumente ausschließlich auf Papier oder einem anderen Beschreibstoff geschriebene Gedanken gelten, d.h. Dokumente sind immer Textdokumente (vgl. z.B. Ranganathan, Shiyali R. (Hrsg.): Documentation and its facets. London 1963). Die Theoriebildung zum Dokument geht dabei meist von der Etymologie des Dokumentbegriffs aus, von seiner Herkunft vom lateinischen *documentum* beziehungsweise dessen Stamm *docere* (= lehren) (vgl. z.B. Lund, Niels Windfeld: „Documentation in a complementary perspective“. In: Rayward, W. Boyd (Hrsg.): Aware and Responsible: Papers of the Nordic-International Colloquium on Social and Cultural Awareness and Responsibility in Library, Information and Documentation Studies (SCARLID). Lanham, Md. 2004, S. 93-102; Pédaque, Roger T.: Le document à la lumière du numérique. Caen 2006, insb. S. 28), das heißt von einer pragmatischen Ebene, einer funktionalistischen Dokumentansicht. Für eine Überblicksdarstellung zu traditionellen Dokumentkonzepten vgl. zum Beispiel Buckland, Michael: „What is a „digital document“?“ <http://people.ischool.berkeley.edu/~buckland/digdoc.html> [Stand 18.12.2011].

7 Briets berühmtes Beispiel ist das der Antilope, die in freier Wildbahn nicht als Dokument aufgefasst werden kann, im Symbolsystem Zoo, in dem sie zum Studienobjekt wird, jedoch sehr wohl. Sie ist in diesem Kontext zum physischen Beleg oder Zeugnis geworden, das zum Studium genutzt wird (vgl. Briet, Suzanne: Qu'est-ce que la documentation? Paris 1951, S. 9ff.).

8 Lund (Anm. 6), S. 99ff.

9 Vgl. z.B. Pédaque (Anm. 6), S. 38.

10 Der Verlust des physischen Trägermediums hat 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. Rein technisch gedacht wird das Verhältnis von Original und Kopie weitestgehend bedeutungslos, sofern nicht die Zeit als zusätzliche Dimension mit hinzugenommen wird (vgl. hierzu Voß, Jakob: „Zur Neubestimmung des Dokumentbegriffs im rein Digitalen“. In: LIBREAS. Library Ideas 2 (2009), <http://www.libreas.eu/ausgabe15/texte/002.htm> [Stand 18.12.2011]). 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 in vielen Aspekten ungelöst sind.

11 Wehde, Susanne: Typographische Kultur: eine zeichentheoretische und kulturgeschichtliche Studie zur Typographie und ihrer Entwicklung. Tübingen 2000, S. 7.

12 Shotton, David: „Semantic Publishing: The coming revolution in scientific journal publishing“. In: Learned Publishing 2 (2009), S. 85-94, DOI: 10.1087/2009202, S. 87.

diese Nachteile in der derzeitigen Übergangsphase mangels etablierter neuer Dokumentkonzepte weitgehend in Kauf genommen.

Das digitale Dokument tritt folglich in eine mimetische Beziehung zum traditionellen Dokument.<sup>13</sup> Dass die Übertragung des analogen Dokumentkonzepts auf das digitale Dokument bislang noch relativ gut funktioniert, verdankt sich dem Umstand, dass digitale Bibliotheken heute noch zu einem großen Teil aus Digitalisaten, also Überführungen von analogen Dokumenten in digitale Signale, bestehen. Aber auch genuin digitale Dokumente werden bislang noch häufig nach dem Vorbild analoger Dokumente gestaltet: als „digital incunabula“.<sup>14</sup> Als bloße Repräsentation des Papierdokuments büßt das digitale Dokument jedoch seine Verlässlichkeit in einer Welt ein, die Kulturtechniken des Zugänglichmachens, Organisierens und Archivierens digitaler Dokumente erst erprobt. Es ist eben nicht mehr das Dokument, für das es sich ausgibt, sondern lediglich eine „trace iconique“.<sup>15</sup> In der beschriebenen Übergangsphase wird auch der Publikationsprozess der Gutenberg-Galaxis zunächst schlicht imitiert: Zwar finden viele der Arbeitsschritte innerhalb der Kreation, Distribution, Evaluation und Organisation von Dokumenten nun in digitaler Umgebung statt, der lineare Publikationsprozess und das Konzept des abgegrenzten Dokuments bleiben jedoch weitgehend intakt.<sup>16</sup>

Diese anfängliche Phase des digitalen Zeitalters hat dementsprechend dem Konzept des Dokuments als monolithischem, klar abgegrenzten Objekt und letztlich auch demjenigen von Bibliothek noch wenig anhaben können. Solange es physische Medien – beispielsweise DVDs, CDs, CD-ROMs – und druckanaloge Formate gibt, können traditionelle Konzepte des Sammelns und Bereitstellens von Medien – die Probleme der Langzeitarchivierung und -verfügbarhaltung einmal außer Acht gelassen – weitestgehend aufrechterhalten werden. Erst die zunehmende Virtualisierung, der Verlust des Mediums, das World Wide Web, neue Austauschformate und Auszeichnungssprachen, Text- bzw. Datamining und semantisch basierte Technologien erschüttern den bekannten Publikationsprozess und das Konzept von Bibliothek – ebenso wie das Konzept von Verlag und wissenschaftlicher Kommunikation. Neue technologische Möglichkeiten, wie beispielsweise die Durchdringung der – druckanalog definierten – Dokumentgrenzen, führen die beschriebene Mimesis jedoch zunehmend ad absurdum: In WWW-basierter Umgebung kann in Bezug auf Publikation, Distribution und Evaluation genuin digitaler Dokumente ein netzartiges

Funktionsmodell entstehen, in dem die Zwangsläufigkeit der linearen Abfolge verloren geht und die Dokumentgrenzen zunehmend verschwimmen.<sup>17</sup> Im Kontext von Linked Open Data ist eine weitere Auflösung des traditionellen Funktionskontinuums bereits zu erkennen.<sup>18</sup>

### 3 These zur Diskretheit

*Das Charakteristikum der Diskretheit digitaler Dokumente verändert den Dokumentbegriff und die Kulturtechniken grundlegend.*

Im Unterschied zu analogen Signalen sind digitale Signale diskret. 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.<sup>19</sup>*

Der Übergang vom analogen zum digitalen Signal stellt insbesondere für Geisteswissenschaften als textbasierte und zu einem großen Teil traditionell hermeneutisch arbeitende Wissenschaften einen grundlegenden Wandel dar. Das semiotische System wird gewechselt: vom alphabetischen zum Binärkode, von analoger Unschärfe zum diskreten Datum. Aus informatischer Perspektive ist digitaler Text ein Bitstream. Als solcher ermöglicht er die Prozessierung von Text und verändert dadurch den Rezeptionsprozess (im weiten Sinne) entscheidend. Texte werden maschinenlesbar und ermöglichen derart – geeignete Austauschformate und Interoperabilität vorausgesetzt – die Bildung großer digitaler Textkorpora. Diese Korpora sind als Daten vielfältig prozessierbar und ermöglichen beispielsweise den Literaturwissenschaften quantitative Methoden der Textuntersuchung, die stylometrische und strukturalistische Analysen – zum Beispiel durch die Anwendung compu-

17 Vgl. Pédaque (Anm. 6); vgl. Gradmann/Meister (Anm. 16).

18 Zur Verwendung des Begriffs „Linked Open Data“ bzw. „Linked Data“ vgl. die dreiteilige Artikelreihe des Hochschulbibliothekszentrums des Landes Nordrhein-Westfalen (hbz): Pohl, Adrian: „Open Data im hbz-Verbund“. In: ProLibris 3 (2010), S. 109–113, [http://www.hbz-nrw.de/dokumentencenter/produkte/iod/aktuell/pohl\\_2010\\_open-data.pdf](http://www.hbz-nrw.de/dokumentencenter/produkte/iod/aktuell/pohl_2010_open-data.pdf) [Stand: 18.12.2011]; ders.; Ostrowski, Felix: „„Linked Data“ – und warum wir uns im hbz-Verbund damit beschäftigen“. In: B.I.T. Online 13/3 (2010), S. 259–268, [http://www.hbz-nrw.de/dokumentencenter/produkte/iod/aktuell/pohl\\_ostrowski\\_2010\\_linked-data.pdf](http://www.hbz-nrw.de/dokumentencenter/produkte/iod/aktuell/pohl_ostrowski_2010_linked-data.pdf) [Stand: 18.12.2011]; dies.: „Zur Konzeption und Implementierung einer Infrastruktur für freie bibliographische Daten“. In: Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis (Hrsg.): Semantic Web & Linked Data. Elemente zukünftiger Informationsinfrastrukturen, Frankfurt am Main, S. 205–216, [http://www.hbz-nrw.de/dokumentencenter/produkte/iod/aktuell/pohl\\_ostrowski\\_2010\\_open-data-infrastruktur.pdf](http://www.hbz-nrw.de/dokumentencenter/produkte/iod/aktuell/pohl_ostrowski_2010_open-data-infrastruktur.pdf) [Stand: 18.12.2011].

19 Kittler, Friedrich: Von der Letter zum Bit . In: Wenzel, Horst; Kittler, Friedrich; Schneider, Manfred (Hrsg.): Gutenberg und die neue Welt, München 1994, S. 105–117, S. 105f.

13 Vgl. McCarty, Willard: Humanities computing. Houndsill u.a. 2005, S. 29.

14 Als „digital incunabula“ bezeichnet Crane „digital objects whose form remains firmly rooted in traditions of print, with HTML and PDF largely mimicking the limitations of their print predecessors“ (Crane, Gregory: „What Do You Do with a Million Books?“ In: D-Lib Magazine 3 (2006), DOI: 10.1045/march2006-crane).

15 Pédaque, Roger T.: La redocumentarisation du monde. Caen 2007, S. 17.

16 Vgl. Gradmann, Stefan; Meister, Jan Christoph: „Digital document and interpretation: re-thinking ‚text‘ and scholarship in electronic settings“. In: Poiesis & Praxis: International Journal of Technology Assessment and Ethics of Science 2 (2008), S. 139–153.

terphilologischer Verfahren auf narratologische Fragestellungen – vereinfachen und in Bezug auf große Textkorpora überhaupt erst ermöglichen. Das bedeutet jedoch auch, dass sich die Grenzen des Dokuments hin zur Codierung, zum Text, zum Inhalt öffnen, wobei betont werden muss, dass dieser Prozess im genannten Kontext die Signifikatenebene nicht verlässt. Allerdings impliziert digitaler Text eine Potenzierung der bereits in der traditionellen Philologie vorliegenden Repräsentationsbedingungen: Digitaler Text ist „Repräsentation einer Repräsentation“, Text wird „zu einem in einer neuen, formaleren Symbolsprache repräsentierten System“.<sup>20</sup> Mit der Umwandlung in alphanumerischen Zeichensatz und letztlich in Binärcodierung wird eine weitere Signifikatenebene, ein weiteres Symbolsystem eingezogen, um Text prozessierbar zu machen. Zwar war es schon unter rein analogen Bedingungen möglich, gewisse Textmerkmale in mühsamer intellektueller Tätigkeit auszuzählen, doch schon die einfache String-Suche, die Suche nach Zeichenketten, in einer digitalen Bibliothek führt durch die schiere Quantität an Text, wie sie etwa über Google Books zur Verfügung steht, zu einer unbekannten neuen Qualität und Kulturtechnik. Die digitalen Texte sind nicht länger durch Buchdeckel abgegrenzte monolithische Einheiten, sondern werden – je nach bereits geleisteter Vorarbeit mit unterschiedlichem Erfolg – in einer maschinellen Analyse zueinander geöffnet. Bibliotheken werden zu digitalen Korpora und als solche zum Ausgangspunkt neuer Rezeptionstechniken. Information Retrieval und quantitative Untersuchungen, die computerlinguistische und statische Verfahren nutzen, basieren auf bearbeiteten Textkorpora und stellen den nächsten aufwändigeren und anspruchsvolleren Schritt dar. In der Computerphilologie, die sich mit computergestützten Methoden literarischen Texten nähert, ist in Abgrenzung zum „close reading“ hermeneutischer Traditionslinien wie des New Criticism der Begriff des „distant reading“ entstanden. Im Gegensatz zum „close reading“, das auf einem sehr kleinen Kanon beruhen muss, kann das „distant reading“ durch Techniken des Textmining größere kultur- und literaturhistorische Linien innerhalb eines sehr großen Textkorpus aufzeigen. „Distant reading“ ist „distanziert“, weil es auf große Korpora statt auf Einzelwerke ausgerichtet ist, statistische Methoden und bibliometrische Analysen statt traditioneller Interpretationsmethoden nutzt. Damit ist ein – wie es Gerhard Lauer zuspricht – „Paradigmenwechsel vom Lesen zum Rechnen der Literatur“ eingeleitet,<sup>21</sup> den der italienische Literaturwissenschaftler Franco Moretti bereits seit Jahren provokant einfordert: „At bottom it's [close reading, Anm. der Verf.] a theological exercise – very solemn treatment of very few texts taken very seriously – whereas what we

really need is a little pact with the devil: we know how to read texts, now let's learn how not to read them.“<sup>22</sup>

Bei all den Möglichkeiten, die digitaler Text aufgrund seiner Diskretheit eröffnet, soll jedoch nicht der Eindruck entstehen, dass im Umgang mit digitalem Text die hermeneutische Ebene durch Maschinen ersetzt werden kann: Erst eine von Menschen gestellte und in einen Algorithmus übersetzte Forschungsfrage liefert mittels Prozessierung großer Datenmengen Ergebnisse, die wiederum der menschlichen Analyse bedürfen. Die Potenziale computergestützter Analysen sind stark vom Granularitätsgrad abhängig, also von der internen Struktur und dem Auszeichnungsgrad digitaler Dokumente, wobei die Auszeichnung, zumindest auf semantischer Ebene, nur in eingeschränktem Maße automatisiert stattfinden kann.

#### 4 These zur Rolle des Textbegriffs

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

Basiert das philologische Verständnis von Text häufig auf semiotischen Theorien – gleichgültig wie eng oder weit der Textbegriff jeweils gefasst ist<sup>23</sup> –, wird das Dokument meist als Materialisierung von Text und als abgegrenzte komplexe semiotische Einheit verstanden.<sup>24</sup> Wenn jedoch beispielsweise Algorithmen Dokumentgrenzen durchdringen oder Hyperlinks Dokumentgrenzen übertreten können, wird der materielle Inhalt des Dokuments, das heißt Daten (im technischen Sinne), wird Text, das heißt die (narrativen) Strukturen (im semiotischen Sinne), definitionskräftiger.

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

22 Moretti, Franco: „Conjectures on World Literature“. In: *New Left Review* 1 (2000), S. 54–68, <http://www.newleftreview.org/?page=article&view=2094> [Stand: 18.12.2011]. Vgl. auch Moretti, Franco: *Kurven, Karten, Stammbäume: Abstrakte Modelle für die Literaturgeschichte*. Frankfurt am Main 2009. Für eine kritische Betrachtung des Informationsparadigmas und statistischer Verfahren in Bezug auf Texte vgl. Jochum, Uwe (Hrsg.): *Das Ende der Bibliothek? Vom Wert des Analogen*. Zeitschrift für Bibliothekswesen und Bibliographie: Sonderband 105. Frankfurt am Main 2011.

23 Vom sehr weiten Textbegriff im Sinne von Kristeva's Verständnis verschiedenster (auch nicht-sprachlicher) kultureller Systeme und Codes als Text (vgl. z.B. Kristeva, Julia: „Bachtin, das Wort, der Dialog und der Roman“. In: *Literaturwissenschaft und Linguistik: Ergebnisse und Perspektiven* 3 (1972), S. 345–375) 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, Gérard: *Palimpseste*: Die Literatur auf zweiter Stufe. 3. Aufl. Frankfurt am Main 2001), bis zur Eingrenzung des Begriffs, die „Textualität an die gesellschaftliche Aufbewahrung von Wissen“ bindet (Knobloch, Clemens: „Text/Textualität“. In: Barck, Karlheinz et al. (Hrsg.): *Ästhetische Grundbegriffe: Historisches Wörterbuch in sieben Bänden*. Stuttgart u.a. 2005, S. 23–48, S. 46).

24 Vgl. z.B. Ørom, Anders: „The concept of information versus the concept of document“. In: Skare, Roswitha; Lund, Niels Windfeld; Varheim, Andreas: *A Document (re)turn: Contributions from a research field in transition*. Frankfurt am Main u.a. 2007, S. 53–72.

20 Meister, Jan Christoph: Projekt Computerphilologie, 2003, [http://edocs.ub.uni-frankfurt.de/volltexte/2008/11645/pdf/MEISTER\\_Projekt.pdf](http://edocs.ub.uni-frankfurt.de/volltexte/2008/11645/pdf/MEISTER_Projekt.pdf) [Stand: 18.12.2011], S. 8.

21 Lauer, Gerhard: „Literatur rechnen: Lektüre im Computerzeitalter“, In: *Faz.net* (26.12.2009), <http://www.faz.net/s/Rub5C2BFD49230B472BA96E0B2CF9FAB88C/Doc~E40438503D67B4942A7FE49FB3324C0BF~ATpl~Ecommon~Scontent.html> [Stand: 18.12.2011].

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)<sup>25</sup> schlicht als „anything that can have an identifier“ definiert.<sup>26</sup> Im WWW verweisen Hyperlinks auf andere Ressourcen und führen auch direkt – um in den im Kontext des WWW üblichen räumlichen Metaphern zu sprechen – an ihren Ort (im Sinne der URL). Hyperlinks stellen im Vergleich zum Zitat, das fremden Text ins Dokument integriert, zur bibliographischen Angabe, die auf einen anderes Dokument zeigt, 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 und verlagern das Lesen an einen anderen Ort.

Auf der Textebene selbst hat die voranschreitende Etablierung von XML (Extensible Markup Language), einer Metasprache, auf deren Basis Auszeichnungssprachen zur Strukturierung digitaler Texte definiert werden können, ein Charakteristikum des traditionellen Dokumentbegriffs fundamental zunichte gemacht: die Einheit von Inhalt und Form, von Beschriftung und Trägermedium.<sup>27</sup> Im XML-Dokument zerfällt diese Einheit 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 hinzu, zielt XML ganz auf die logische Struktur von Dokumenten ab. Im Gegensatz zu HTML-Dokumenten sind XML-Dokumente dadurch nicht nur für Menschen, sondern auch für Maschinen lesbar, das heißt prozessierbar – und somit derzeit das tragende Fundament von Linked Data.<sup>28</sup>

Die Fokussierung auf den – diskreten – Inhalt von Dokumenten, geht im Kontext der Entwicklung semantisch basierter Technologien mit der Hoffnung einher, über die Repräsentation des Inhalts beispielsweise eine bessere Auffindbarkeit von Information zu erreichen, die Beantwortung komplexer Suchanfragen durch automatisierte Auswertung der Informationen im Netz oder sogar die Gewinnung weiterer Information durch automatische Schlussfolgerungen aus maschinenlesbarer Information zu ermöglichen. Es ist dementsprechend naheliegend, dass das W3C das „web of documents“ langfristig durch ein „web of data“ abzulösen beziehungsweise dem „web of documents“ ein „web of data“ hinzuzufügen gedenkt.<sup>29</sup> Berners-Lee erklärte bereits

im Jahr 2001: „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“.<sup>30</sup> Allerdings ermöglichen erst Wissensorganisations-Systeme den Einsatz semantisch basierter Technologien. Dabei liegt es nahe, dass Bibliotheken nicht nur ihre bibliographischen Daten in die Linked-Open-Data-Sphäre einbringen, sondern als traditionelle Akteure im Bildungssystem und Wissensmanagement auch Wege in die Informationsfülle des WWW ebnen. Anders als bei klassischen Dokumentationssprachen und Klassifikationen sehen sich Bibliotheken dabei jedoch nicht mehr mit eigenen Konzeptionen, sondern mit solchen der Informatik konfrontiert.

Auf Grundlage des Resource Description Framework (RDF), einer Beschreibungssprache für WWW-Ressourcen, und RDF Schema (RDFS), das als Grammatik von RDF fungiert, lassen sich bereits vielfältige Beziehungen zwischen Ressourcen im WWW herstellen. Ontologiesprachen wie OWL (Web Ontology Language) stellen darauf basierend komplexere Wissensrepräsentations-Systeme her. Auffällig ist in diesem Kontext, dass von informatischer Seite Dokumente häufig mit einer Signifikantenebene identifiziert werden, von der das „web of data“ befreit erscheint: „[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*“.<sup>31</sup> Dass durch Linked Data statt Dokumenten die „wirklichen“ Dinge in der Welt beziehungsweise die Worte, die auf die Dinge der Welt verweisen, und vermeintlich universelle Konzepte verbunden werden können, basiert auf einer „(technically) high-level nominalist regression“,<sup>32</sup> die jegliches semiotische Modell ignoriert und vor allem Text als eigenständiges semiotisches System unberücksichtigt lässt. Problematisch sind dabei nicht in erster Linie die Ontologien selbst, sondern das „ontology-mapping“ beziehungsweise die „data fusion“, also der Prozess der Integration unterschiedlicher Dateneinheiten zu einer „single, consistent, and clean representation“,<sup>33</sup> das Van Harmelen zu Recht als „the Achilles Heel of the Semantic Web“ bezeichnet.<sup>34</sup> Zu hinterfragen ist, ob jenseits kontrollierter fachwissenschaftlicher Vokabulare „(essentially non-lingual) Unit of Thought“, die „concepts“, existieren, die lediglich unterschiedliche „identifiers“ oder „terms“ haben

25 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.

26 Booth, David; Haas, Hugo et al.: Web Services Architecture: W3C Working Group Note 11 February 2004, <http://www.w3.org/TR/ws-arch/> [Stand: 18.12.2011].

27 Vgl. Pédaque (Anm. 6).

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

29 Vgl. z.B. W3C: Semantic Web, 2010, <http://www.w3.org/standards/semanticweb/> [Stand: 18.12.2011].

30 Berners-Lee, Tim; Hendler, James; Lassila, Ora: „The Semantic Web“. In: *Scientific American* 284 (2001), S. 34-43, S. 37.

31 Sauermann, Leo; Cyganiak, Richard: Cool URIs for the Semantic Web: W3C Interest Group Note 03 December 2008, <http://www.w3.org/TR/cooluris/> [Stand: 18.12.2011].

32 Gradmann/Meister (Anm. 16), S. 147.

33 Bizer, Christian; Heath, Tom; Berners-Lee, Tim: „Linked Data – The Story So Far“. In: *International Journal on Semantic Web and Information Systems* 5 (2009), S. 1-22, <http://eprints.ecs.soton.ac.uk/21285/1/bizer-heath-berners-lee-ijsws-linked-data.pdf> [Stand: 18.12.2011].

34 Van Harmelen, Frank: Semantic Web Research Anno 2006: Main Streams, Popular Fallacies, Current Status and Future Challenges. Berlin/Heidelberg 2006.

können.<sup>35</sup> Die Vorstellung von festen denotativen Bedeutungsbeziehungen, das Ignorieren der Kontextabhängigkeit von Sprache im Sinne Wittgensteinscher Sprachspiele, der Unübersetbarkeit kultureller Konzepte, des „Kreis[es] der unbegrenzten Semiose“,<sup>36</sup> der Problematik abgeschlossener Semantiken, fester Klassifikationen, statischer „Ontologien“ – kurz: das Ignorieren der gesamten Traditionslinie von Nietzsches Sprachkritik über die Frankfurter Schule bis zum Poststrukturalismus – lassen diesen Ansatz verkürzt und problematisch erscheinen.<sup>37</sup>

Interessant ist die mit Linked Open Data entstehende Möglichkeit zur – im Vergleich mit Hyperlinks – differenzierteren Vernetzung von Daten, zum Beispiel Texten. Der Begriff des Dokuments, das in RDF-Tripel zergliedert wird, tritt dabei zunächst zugunsten seines Inhalts und vor allem kleinerer Informationseinheiten des Inhalts in den Hintergrund. Auch im bibliothekarischen Bereich, in dem in zunehmendem Maße ein elektronischer oder zumindest hybrider Bestand verwaltet werden muss und bibliographische Daten potenziell auch für die Nachnutzung durch semantisch basierte Anwendungen modelliert werden sollen, zeigt sich das Umdenken vom Dokument zum Inhalt im digitalen Paradigma – wobei an einem fast platonisch anmutenden Werkbegriff festgehalten wird, der fragwürdig erscheint. 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 – beispielsweise im Kontext von Linked Data – genutzt werden. Daher wandert der Fokus von der Titelaufnahme, die auf ein Dokument verweist, zum Inhalt, zu den „things‘ the data describe“.<sup>38</sup> Die RDA werden gegenüber ihrem Vorläufermodell, der zweiten Ausgabe der Anglo-American Cataloguing Rules (AACR2), explizit als „designed for the digital world“ herausgestellt – wahrscheinlich wäre zutreffender: für eine hybride Welt.<sup>39</sup> Waren

35 Mons, Barend; Velterop, Jan: „Nano-Publication in the E-Science Era“. In: Proceedings of the Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009), <http://www.surfoundation.nl/SiteCollectionDocuments/Nano-Publication%20-%20Mons%20-%20Velterop.pdf> [Stand: 18.12.2011].

36 Eco, Umberto: Zeichen: Einführung in einen Begriff und seine Geschichte. 1. Aufl. Frankfurt am Main 1977, S. 179f.

37 Fraglich ist zudem 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. Einen Schritt in diese Richtung geht Calais ([www.opencalais.com](http://www.opencalais.com)): Der Calais Web Service wandelt unstrukturierte Dokumente in RDF um. Darüber hinaus ist bislang nicht abzusehen, wie sich in sehr großen Triplestores Performance und Skalierbarkeit gestalten.

38 IFLA Study Group on the Functional Requirements for Bibliographic Records (Hrsg.): Functional Requirements for Bibliographic Records, 1998, <http://archive.ifla.org/VII/s13/frbr/frbr1.htm> [Stand: 18.12.2011].

39 Vgl. Joint Steering Committee for Development of RDA: „RDA – Resource Description and Access: A Prospectus“ (01.07.2009), <http://www.rda-jsc.org/rdaprospectus.html> [Stand: 18.12.2011].

früher die jeweilige Ausgabe des Werkes und die davon verfügbaren Exemplare, also die Dokumente, zentral, wird nun der Ausgang beim Originalwerk als abstrakter intellektueller Einheit genommen: Es wird zwischen dem Inhalt und seiner Darstellung unterschieden, was sich beispielsweise daran zeigt, dass mit einem Werk mehrere Expressionen<sup>40</sup> und Manifestationen<sup>41</sup> verknüpft werden können.

Bei all diesen Beobachtungen der Auflösung des Dokuments und des Dokumentbegriffs ist bislang noch weitgehend ungeklärt, wie der Inhalt, wie der Text, der nun ohne klare Grenzen in vernetzter Umgebung vorliegt, die pragmatisch-gesellschaftlichen Funktionen des Dokuments übernehmen kann.

## 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.*

Wie alle Technologien ist auch Informationstechnologie nicht neutral: Sie gibt Möglichkeiten und Grenzen des Handelns und Kommunizierens vor. „Program or be programmed“ ist die griffige Formel, mit der amerikanische Kolumnist und Hochschullehrer Douglas Rushkoff digitale Mündigkeit einfordert: „A New Medium Requires a New Literacy“.<sup>42</sup> Vilém Flusser bemerkte schon im letzten Jahrtausend: „Wer die neuen Codes nicht lesen kann, ist Analphabet in einem mindestens so radikalen Sinn, wie es die der Schrift Unkundigen in der Vergangenheit waren“.<sup>43</sup> Häufig scheint jedoch in Vergessenheit zu geraten, dass Informationstechnologie selbst historisch entstanden ist und vorherrschende Diskurse und Denkweisen abbildet, die in der Folge wiederum den Umgang mit Information, Text und Dokument bestimmen.<sup>44</sup>

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 *episteme*, als Bedingung der Möglichkeit des Wissens in der westlichen

40 Z.B. eine französische Originalfassung und eine deutsche Übersetzung und eine überarbeitete französische Fassung aus einem späteren Jahr.

41 Z.B. eine Buchausgabe, eine PDF-Ausgabe, eine HTML-Ausgabe.

42 Rushkoff, Douglas: „A New Medium Requires a New Literacy“. In: Focus Online (25.10.2010), [http://www.focus.de/digital/internet/dld-2011/debate-english/tid-20544/douglas-rushkoff-a-new-medium-requires-a-new-literacy\\_aid\\_575297.html](http://www.focus.de/digital/internet/dld-2011/debate-english/tid-20544/douglas-rushkoff-a-new-medium-requires-a-new-literacy_aid_575297.html) [Stand: 18.12.2011] bzw. Rushkoff, Douglas: Program or Be Programmed: Ten Commands for a Digital Age. New York: OR Books, 2010.

43 Flusser (Anm. 1), S. 50.

44 Pédaque betonen daher: „il est nécessaire de quitter toute conception neutraliste de la technique“ (Pédaque (Anm. 6), S. 96). Und in den Digital Humanities wird gefordert: „Like the historians of computing, digital scholars must learn to read software to elicit the history and practice that it embodies“ (Mahoney, Michael S.: „The Histories of Computing(s)“. In: Interdisciplinary Science Reviews 30 (2005), S. 119–135, DOI: 10.1179/030801805X25927, S. 132).

Welt, die auch die Informationstechnologie strukturiert.<sup>45</sup> Auf der Ebene der Markup-Sprachen zeigt sie sich unter anderem im Aufbau von XML: XML-Sprachen erfordern, dass ein Dokument stets „wohlgeformt“ zu sein hat, das heißt, dass alle Elemente geöffnet und geschlossen werden und sauber geschachtelt sein müssen, es darf keine Überlappungen und nur ein Wurzelement geben. Die Wohlgeformtheit eines XML-Dokuments wird von XML-Parsern überwacht, von Programmen, die ein XML-Dokument einlesen, überprüfen und verarbeiten. Für Ted Nelson, der mit Xanadu ein alternatives System schaffen wollte,<sup>46</sup> ist XML daher „not an improvement but a hierarchy hamburger: Everything, everything must be forced into hierarchical templates!“<sup>47</sup>

Auch den unterschiedlichen Formaten liegen Daten- und Textmodelle zugrunde. Renear, einer der Protagonisten im Kontext der Theoriebildung um die Text Encoding Initiative (TEI),<sup>48</sup> kategorisiert Text wie folgt:

45 Im Bibliotheksgebiet spiegelt sich diese *episteme* in hierarchischen Klassifikationen wider. Für Nelson stellt die „simulation of hierarchy“ die „principal tradition in the computer world“ dar (Nelson, Theodor Holm: „Toward A Deep Electronic Literature: The Generalization of Documents and Media“, 2007, <http://xanadu.com/XanaduSpace/xuGzn.htm> [Stand: 18.12.2011]). 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 (Anm. 13), S. 195).

46 Xanadu ist eine Software, die einen andersgearteten Umgang mit Dokumenten ermöglichen soll als er im heutigen WWW realisiert ist. Nelson arbeitet seit den 60er Jahren des vorigen Jahrhunderts an Xanadu – allerdings ist bis heute nur eine äußerst minimalistische Demo-Version herunterzuladen. Zur Geschichte des Xanadu-Projekts vgl. Wolf, Gary: „The Curse of Xanadu“, 1995, <http://www.signallake.com/innovation/CurseOfXanaduJun95.pdf> [Stand: 18.12.2011].

47 Nelson, Theodor Holm: „I Don't Buy In“. <http://ted.hyperland.com/buyin.txt> [Stand: 18.12.2011].

48 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 Codierung 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, Lou; Bauman, Syd (Hrsg.): TEI P5: Guidelines for Electronic Text Encoding and Interchange, 2009, <http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf> [Stand: 18.12.2011], 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. 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 (Anm. 13), 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, Fotis: „Computerphilologie“. In: Anz, Thomas (Hrsg.): Handbuch Literaturwissenschaft: Gegenstände–Konzepte–Institutionen. Bd. 2., Methoden und Theo-

*one could model text as a sequence of graphic characters, as in the „plain vanilla ASCII“ approach of Project Gutenberg; as a combination of procedural 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.<sup>49</sup>*

Abgesehen von den Bildrepräsentationen von Text, die aufgrund ihrer eingeschränkten Funktionalitäten zu vernachlässigen sind, ist Text vom informatischen Standpunkt betrachtet eine Folge alphanumerischer Zeichen. Das Datenmodell von Text ist mit literaturwissenschaftlichen Textmodellen in dieser Form jedoch nicht kompatibel. Buzzetti zufolge ist die Struktur der digitalen Repräsentation entscheidend, wenn es um die Einschätzung adäquater Bedingungen für den Umgang mit Texten geht, weil erst sie die automatische Prozessierung des Informationsgehalts eines Dokuments ermöglicht. 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 (wenn auch nicht ausschließlich) in der grafischen Gestaltung des Textes sichtbar wird.<sup>50</sup>

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. Als den genannten Codierungsarten überlegen hat für Renear digitaler Text zu gelten, der nach den TEI-Richtlinien in XML (damals noch SGML) ausgezeichnet ist. TEI-getaggte Texte basieren Renear zufolge auf dem Modell von Text als einer „ordered hierarchy of content objects“ (OHCO): Text sei hierarchisch aufgebaut, die ihn konstituierenden Objekte

rien, Stuttgart 2007, S. 27-40, 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, Susan: „The Text Encoding Initiative. An Interchange Format Once Again“, 2009, <http://computerphilologie.tu-darmstadt.de/jg08/schreibman.html> [Stand: 18.12.2011]). Diese Anpassung wird zwar in den verschiedenen Versionen der Guidelines vollzogen, allerdings stellen die TEI Guidelines schon allein aufgrund ihres Umfangs kein besonders flexibles System für Bedeutungswandel dar.

49 Renear, Allen: „Text Encoding“. In: Schreibman, Susan; Siemens, Raymond George; Unsworth, John (Hrsg.): A companion to digital humanities. Malden, Mass. 2004, S. 218-237, S. 225.

50 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 als PDF vorliegenden Artikeln automatisch in XML (genauer in TEI) encodieren kann – und zwar, indem sie grafische Muster der PDF-Vorlage auswertet. Vgl. Lopez, Patrice: „GROBID: Combining Automatic Bibliographic Data Recognition and Term Extraction for Scholarship Publications“, In: Proceedings of ECDL 2009, <http://scite-it.eu:8080/grobid/doc/final.pdf> [Stand: 18.12.2011] und die Betaversion, zu finden unter: <http://grobid.no-ip.org> [Stand: 18.12.2011].

(zum Beispiel: Kapitel, Absatz) ineinander geschachtelt.<sup>51</sup> Interessant ist in diesem Kontext, dass Renears Modell tatsächlich aus der Praxis der Textcodierung erwachsen 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 (Buch-)Repräsentation) oder bei der Encodierung von Lyrik stellt.<sup>52</sup> „Content objects“ wie Sätze und (einem bestimmten Metrum folgende) Verse in Gedichten sind kaum in einer Hierarchie unterzubringen, Enjambements und Verse wiederum nicht. Renear schwächt daher sein OHCO-Modell ab: „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“.<sup>53</sup> Allerdings wird Renears OHCO-Modell auch in seiner abgeschwächten Form in Frage gestellt – häufig mit Bezug auf die Encodierung von Lyrik,<sup>54</sup> denn überlappende Strukturen und das Prinzip der Wiederholung (z.B. Reim, Metrik, phonetische Stilfiguren) stellen die Essenz von Lyrik dar und zeichnen auch andere literarische Gattungen aus. So betont McGann, dass literarische Texte nicht nur nach ihren linguistischen Elementen zu beurteilen seien, sondern nach den Beziehungen, die ihr semiotisches Material bildeten. In der linguistischen Hierarchie, die TEI und SGML/XML widerspiegeln, würde Text als ein Container für Information verstanden. Diese Sicht sei jedoch mit rhetorischen Strukturen von Sprache nicht vereinbar. McGann betont, dass sich das in Bergen angesiedelte Wittgenstein-Editionsprojekt nicht ohne Grund gegen TEI/SGML als Auszeichnungssystem der Wittgensteinschen Texte entschieden habe.<sup>55</sup> Kants

51 DeRose, Steven J.; Durand, David G.; Mylonas, Elli; Renear, Allen H.: „What Is Text, Really?“, In: Journal of Computing in Higher Education 1 (1990), S. 3-26, <http://antonietta.philo.unibo.it/IUcorso2006-07/risorse/ohco.pdf> [Stand: 18.12.2011].

52 In den TEI Guidelines P5 gibt es einige Vorschläge zum Umgang mit der „inability to encode overlapping hierarchies“ (Burnard/Bauman (Anm. 48), S. 634).

53 Renear, Allen; Mylonas, Elli; Durand, David: „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> [Stand: 18.12.2011], S. 11.

54 Beispielsweise auf einer Konferenz der Association for Computers and the Humanities/Association for Literary and Linguistic Computing (ACH/ALLC) im Jahr 1999: Dort betont McGann, dass sich insbesondere Lyrik dem OHCO-Modell entzieht. McGann wiederholt diese Position auch bei späteren Gelegenheiten: „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, Geoffrey: „What is Text Analysis, Really?“ In: Literary and Linguistic Computing 18 (2003), S. 209-219, S. 209.)

55 Renear, Allen; McGann, Jerome; Hockey, Susan: „Abstract for ACHALLC99 debate: What is Text?“, 1999, [http://www.humanities.ualberta.ca/Susan\\_Hockey/achallc99.htm](http://www.humanities.ualberta.ca/Susan_Hockey/achallc99.htm) [zurzeit nicht verfügbar; Stand: 18.12.2011]. Darüber hinaus sind derartige Textmodelle unvereinbar mit literaturwissenschaftlichen Richtungen, die die Materialität

Texte funktionieren nach anderen Mechanismen als Texte Derridas‘ – je nach Schreibweise des jeweiligen Autors, die eher logisch oder assoziativ sein kann, eignet sich eine hierarchische Auszeichnung – in diesem Kontext stets als anspruchsvolle detaillierte Auszeichnung verstanden – mehr oder weniger gut.

Die dargestellte Debatte vermag einige Aspekte hinsichtlich Codierungsarten, Formaten und Textmodellen aufzuzeigen: Jeder technischen Entscheidung liegt (bewusst oder unbewusst) ein (Text-) Modell zugrunde, dessen Konsistenz erst die Praxis zeigt. Häufig erweisen sich unterschiedliche Perspektiven auf ein Phänomen – hier die logisch-linguistische mit einer semiotisch-semantischen – als miteinander unvereinbar. Dabei ist anzumerken, dass die Auszeichnung literarischer Texte nach TEI in wissenschaftlichen Editionsprojekten noch immer alternativlos ist – übrigens führt Renear diese Tatsache stets als ein Argument zur Verifizierung des OHCO-Modells an.<sup>56</sup>

Das Problem der überlappenden Hierarchien lässt sich jedoch auch aus einer anderen Perspektive betrachten: im Kontext der Tradition des Kommentars, denn in der Rezeptionsgeschichte eines literarischen, philosophischen oder religiösen Textes schichten sich häufig miteinander nicht kompatible Interpretationen über- und ineinander.<sup>57</sup> Markup stellt immer eine Interpretation des Textes dar. Im Vorwort zu den den TEI-Richtlinien klingt das folgendermaßen:

*Generalizing from that sense, we define markup [...] 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.<sup>58</sup>*

des Textes ins Zentrum ihrer Untersuchung rücken. Es berücksichtigt nicht 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 (Anm. 11), S. 11).

56 Vgl. z.B. Renear (Anm. 49), S. 225.

57 Dieser Prozess ist mit demjenigen vergleichbar, den Benjamin in seinem Lesskow-Aufsatz mit oralen Erzähltraditionen identifiziert: „jenes langsame Einander-Überdecken dünner und transparenter Schichten“, die „Schichtung vielfacher Nacherzählungen“ (Benjamin, Walter: „Der Erzähler. Betrachtungen zum Werk Nikolai Lesskows“. In: ders.: Gesammelte Schriften. Frankfurt am Main 1991, Bd. II/2, S. 438-465, S. 448).

58 Burnard/Bauman (Anm. 48), S. xxxi.

Dass es angesichts der Vielzahl an möglichen Interpretationen „multiple hierarchies“<sup>59</sup> gibt, verwundert kaum. Wenn Markup im Sinne des Explizitmachens des im Druckmedium Impliziten, aber auch im Sinne semantischer Auszeichnungen den Text begleitet, entsteht mit seinen unterschiedlichen Markup-Versionen eine Interpretations- und Rezeptionsgeschichte des Textes. Die Unschärfe des Textes verschwindet nicht durch seine diskrete Codierung in digitalen Signalen, aber Kulturtechniken der Rezeption, die in Bezug auf gedruckte Texte lediglich typographisch zum Ausdruck kamen und ansonsten als gesellschaftliche Prägung die Rezeption eines Textes strukturierten, werden durch TEI-Auszeichnung festgeschrieben. Darauf hinaus wandert in digitalen Editionen literarischer Texte der Kommentar in den Primärtext beziehungsweise kann mit dem Primärtext überblendet werden. Diese Encodierungen, diese Kommentare treten als „a reconfiguration which advances knowledge through a discourse realisable only within an electronic environment“ auf.<sup>60</sup> Zusammenfassend lässt sich konstatieren, dass die reine Binärcodierung von Texten einen simplifizierenden Textbegriff befördert, der durch Markup in Analogie zum überlieferten Textbegriff gebracht wird.

## 6 These zur Rekonstruktion des Dokumentbegriffs

*Im Zuge einer Rekonstruktion des Dokumentbegriffs unter digitalen Bedingungen entstehen neuartige digitale Dokumente und Kulturtechniken, auf die sich Bibliotheken einstellen müssen und an deren Stabilisierung sie mitwirken können.*

Roger T. Pédaque, ein interdisziplinäres französisches Autorenkollektiv, sieht 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.<sup>61</sup> Absicht der „documentarisation“ war es, den Nutzen von Dokumenten, den Zugang zu ihren Inhalten zu verbessern: „on peut dire que documentariser, c'est donner un statut à un texte, le faire „document“.“<sup>62</sup> Ein ähnliches Streben, wenn auch unter veränderten Bedingungen und

59 Vgl. Witt, Andreas: „Multiple Hierarchies: New Aspects of an Old Solution“, 2006, <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.135.2648&rep=rep1&type=pdf> [Stand: 18.12.2011].

60 Schreibman, Susan: „Computer-Mediated Texts and Textuality: Theory and Practice“. In: Computers and the Humanities 36 (2002), S. 283–293, S. 287.

61 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: **Réseau Thématique Pluridisciplinaire „Document et contenus : création, indexation, navigation“**. Beteiligt waren über 100 Wissenschaftlerinnen und Wissenschaftler, u.a. aus den Disziplinen Informatik, Linguistik, Philosophie, Soziologie, Mathematik, Bibliotheks- und Informationswissenschaft. Vgl. Pédaque (Anm. 6); Pédaque (Anm. 15).

62 Salaün, Jean-Michel: „Web, texte, conversation et redocumentarisation“. In: Actes des 9èmes journées internationales d’analyse statistique des données textuelles. Lyon 2008, S. 27–30, <http://hdl.handle.net/1866/2226>, S. 6.

mit anderen Mitteln, ist im digitalen Zeitalter nach der anfänglichen Emulationsphase zu beobachten.<sup>63</sup> 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 und die Möglichkeiten der digitalen Umgebung ungenutzt zu lassen.

Als Beispiel dafür können schon heute digitale Editionen angeführt werden. 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 – Konstruertheit des literarischen Textes sichtbar zu machen, muss großen digitalen Editionsprojekten Anreiz gewesen sein. Konnten Faksimile-Editionen und lange Anmerkungsapparate zu historisch-kritischen Ausgaben nur ein Neben- oder (zeitliches) Hintereinander und nur unzureichend einen Wechsel zwischen vielen verschiedenen Fassungen darstellen, ist mit synoptischen Editionen im Digitalen nun technisch realisierbar, worauf manche Buchedition hinzustreben schien.

Mit den neuen technischen Möglichkeiten gehen auch neue Arbeitstechniken und ein neues Selbstverständnis der Geisteswissenschaften einher. Obwohl die Beobachtung, dass sich mit der Durchsetzung neuer Technologien neue künstlerische (und wissenschaftliche) Formen verbreiten, zutrifft, lässt sich die Argumentation durchaus umkehren: „Die Geschichte jeder Kunstform hat kritische Zeiten, in denen diese Form auf Effekte hindrägt, die sich zwangsläufig erst bei einem veränderten technischen Standard, d.h. in einer neuen Kunstform ergeben können“, heißt es bei Walter Benjamin.<sup>64</sup> Unter digitalen Bedingungen hat sich für das wissenschaftliche Arbeiten ein ganzer Apparat von Methoden und Werkzeugen ausgebildet, der im Rahmen virtueller Forschungsumgebungen zur Verfügung gestellt wird und der einer besseren Nutzung von Dokumenten und digitalen Korpora sowie dem kollaborativen Arbeiten dient.<sup>65</sup> Die in den letzten Jahren zu beobachtende Tendenz im Bereich des naturwissenschaftlichen elektronischen Publizierens, in Prosaform verfasste Artikel in RDF-Strukturen aufzulösen – in „Nano-Publications“<sup>66</sup> – bzw. auch Primärdaten als Publikation zu verstehen, ist auf den geisteswissenschaftlichen Bereich in ähnlicher Form schwer übertragbar. Das ist vor allem dem Umstand geschuldet, dass Geisteswissenschaften,

63 Für eine tabellarische Vergleich zwischen „documentarisation“ und „redocumentarisation“ vgl. Salaün (Anm. 62), S. 8.

64 Benjamin, Walter: „Das Kunstwerk im Zeitalter seiner technischen Reproduzierbarkeit“. In: ders: Gesammelte Schriften. Frankfurt am Main 1991, Bd. I/2, S. 431–469, S. 462f.

65 So bemüht sich die Deutsche Forschungsgemeinschaft durch die Förderung von Textgrid Geisteswissenschaftlerinnen und -wissenschaftlern digitale Werkzeuge wie XML-Editoren, Lemmatizer u.Ä. zur Verfügung zu stellen.

66 Vgl. Mons/Velterop (Anm. 35); vgl. Groth, Paul; Gibson, Andrew; Velterop, Johannes: „The Anatomy of a Nanopublication“. In: Information Services and Use 30 (2010), S. 51–56, <http://www.w3.org/wiki/images/c/c0/HCLSIG%24SWANSIOC%24Actions%24RhetoricalStructure%24meetings%2420100215%24cwa-anatomy-nanopub-v3.pdf> [Stand: 18.12.2011].

die nicht empirisch arbeiten, anders als Naturwissenschaften keine klar definierten Datensets besitzen. Dennoch ist in Kunst wie in Wissenschaft unter digitalen Bedingungen eine Entwicklung von der großen Prosaform – sei es dem Roman oder der Monographie als Werk eines Autors oder einer Autorin – hin zu kleineren, zergliederbaren Formen festzustellen.<sup>67</sup> Das bedeutet beispielsweise im literarischen Bereich nicht, dass Formen einander ablösen: „Die Malerei ist nicht tot. Der Roman ist nicht tot. Sie sind nur nicht mehr von so zentraler kultureller Bedeutung, wie sie es einst waren.“<sup>68</sup> Auf dem amerikanischen E-Book-Markt entstehen mit „Singles“ kürzere Formen, und der amerikanische Schriftsteller David Shields fordert in seinem Manifest „Reality Hunger“: „Wir müssen uns allmählich mal darum bemühen herauszufinden, wie man eine Geschichte fürs Handy erzählt. Eines aber weiß ich: Es ist etwas ganz anderes, als wenn man eine Geschichte für eine DVD mit voller Spielzeit erzählt.“ Shields konstatiert auch, dass Bücher „ihre schwindende Autorität in unserer Kultur nur dann behalten oder wiedererlangen, wenn sie Texte in diese Bibliothek [die digitale Universalbibliothek, Anm. der Verf.] einspeisen“ und „die Suchtechnik aus isolierten Büchern eine Universalbibliothek des gesamten menschlichen Wissens“ macht.<sup>69</sup> Stehen Bücher als physisch isolierte Entitäten im Regal, entwickelt digitaler Text Synapsen. Für Kevin Kelly sind Links und Tags gar „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.*<sup>70</sup>

Tatsächlich eröffnet sich insbesondere im Kontext von Linked Data die Möglichkeit einer differenzierteren Kontextualisierung von (digitalen) Dokumenten und bibliographischen Metadaten – und damit der Nachnutzung bereits vorhandener genuin bibliothekarischer Ressourcen.<sup>71</sup> Die

Frage, die sich in dieser Situation stellt, zielt zugleich ins Herz von Pédaques „redocumentarisation“-Bemühungen: Wie ist ein stabiles digitales Dokument zu schaffen, ohne PDF-ähnliche Techniken zu nutzen, die den dynamischen Charakter des WWW konterkarieren? Einen Weg zeigt die Open Archives Initiative Object Reuse and Exchange (OAI-ORE) auf, die auf der Basis von RDF Standards für die Beschreibung und den Austausch von Ressource-Aggregationen definiert.<sup>72</sup> Erwachsen ist OAI-ORE eben aus dem Problem bewusstsein, dass es keinen Standard zur Identifizierung der derzeit im WWW sehr ephemeren vorliegenden Aggregationen gibt.<sup>73</sup> Wenn jedoch definierte RDF-Strukturen in Form einer Aggregation stabilisiert werden, könnte diese Aggregation dokumentähnliche Funktionen mit neuen Funktionalitäten übernehmen. Indem eine URI für die Aggregation vergeben und zusätzlich eine „Resource Map“ verzeichnet wird, die eine eigene URI hat und die – für Maschinen und Menschen lesbar – zur Beschreibung der Aggregation dient, ist zugleich eine Grenze gezogen und unterschieden, welche Ressourcen Teil der Aggregation und welche ihr nur verbunden sind.

Aggregationen könnten beispielsweise verschiedene Versionen eines Textes beinhalten. Im Bereich der Naturwissenschaften verdrängen Publikationen von Primärdaten zusammen mit Prozessierungsmethoden bereits konventionell in Textform verfasste Publikationen. Sind bei empirischen Wissenschaften Primärdaten beispielsweise Messwerte aus Versuchsreihen, können in den textbasierten Geisteswissenschaften nur wiederum Texte als (wenig granulare) Primärdaten, als ihre Quellen, aufgefasst werden: im Bereich der Philologien beispielsweise die Primärtexte, ohne die ein Verständnis der Sekundärliteratur zumeist kaum 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.<sup>74</sup> Als Aggregation verschiedenartig getagter Texte erhielte man so eine ganze Rezeptionsgeschichte des Primärtextes.

In Ansätzen des Semantic Publishing werden derzeit neue Funktionalitäten derartiger digitaler Dokumente, so genannter „enhanced publications“, erprobt: „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 integration between papers“.<sup>75</sup> Derart veränderte Dokumente stellen eine neue

67 Zu den Veränderungen des wissenschaftlichen Arbeitens und der wissenschaftlichen Prosa vgl. Hayles, N. Katherine: „How We Think: The Transforming Power of Digital Technologies“ (15.01.2009) [Vortrag], <http://hdl.handle.net/1853/27680>.

68 Shields, David: Reality Hunger: Ein Manifest. München 2011, S. 28.

69 Shields (Anm. 68), S. 36f.

70 Kelly, Kevin: „Scan This Book!“. In: The New York Times (14.05.2006), [http://www.nytimes.com/2006/05/14/magazine/14publishing.html?\\_r=1](http://www.nytimes.com/2006/05/14/magazine/14publishing.html?_r=1) [Stand: 18.12.2011].

71 Vgl. z.B. die RDF-Veröffentlichung der Normdaten der Deutschen Nationalbibliothek zur Nachnutzung im Kontext von Linked Data, die „Bibliographic ontology“, ein RDF-Vokabular für bibliographische Daten (vgl. Voß, Jakob: „Die Bibliographic Ontology als Nachfolger bibliographischer Datenformate?“ (25.11.2009), <http://www.slideshare.net/nichtich/die-bibliographic-ontology-als-nachfolger-bibliographischer-datenformate> [Stand: 18.12.2011]), lobid.org, den Linked-Open-Data-Dienst des hbz zur Konvertierung existierender bibliographischer Daten zu Linked Open Data, und culturegraph.org, einen Linked-Open-Data-Service von hbz und DNB, der dazu dient, gemeinsame Identifikatoren für kulturelle Erzeugnisse (z.B. Texte) zu verwenden.

72 Für einen einführenden Überblick vgl. Gradmann, Stefan; Horstmann, Wolfram: „OAI Object Reuse and Exchange (ORE)“. In: Technology-Watch-Report (14.12.2010), <http://metadaten-twr.org/2010/12/14/oai-ore> [Stand: 18.12.2011].

73 Lagoze, Carl; Van de Sompel, Herbert: „ORE User Guide – Primer“. In: ORE Specifications and User Guides (17.10.2008), <http://www.openarchives.org/ore/1.0/> [Stand: 18.12.2011].

74 Vgl. Genette (Anm. 23).

75 Shotton, David: „Semantic Publishing: The coming revolution in scientific journal publishing“. In: Learned Publishing Nummer 2

Herausforderung für Bibliotheken dar, die sich traditionell darum bemühen, Zugangsmöglichkeiten zu Inhalten und Informationen zu bieten und zudem deren Archivierung zu gewährleisten. Zusätzlich bedeutet die zunehmende Bedeutungslosigkeit des Konzepts von Kopie eine einschneidende Veränderung im traditionellen Denken in physischen Einheiten, den Exemplaren; denn prinzipiell können digitale Dokumente – derzeitige Geschäftsmodelle einmal außer Acht gelassen – zentral gehostet und zur Verfügung gestellt werden.<sup>76</sup> Traditionelle Geschäftsgänge sind angesichts der

(2009), S. 85-94, DOI: 10.1087/2009202, S. 86. 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.

<sup>76</sup> Vgl. hierzu etwa die Nationale Hosting-Strategie im Kontext der Schwerpunkt-Initiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen ([http://www.allianzinitiative.de/de/handlungsfelder/nationale\\_hosting\\_strategie/](http://www.allianzinitiative.de/de/handlungsfelder/nationale_hosting_strategie/) [Stand: 18.12.2011]).

neuen Dokumente dabei kaum aufrechtzuerhalten. Trotzdem können Bibliotheken ihre Funktion des Zugänglichmachens von Dokumenten auch im Digitalen leisten, indem sie einen lokalen, zielgruppengerechten – auf die Universität, die Forschungsgemeinde, die Stadt zugeschnittenen – Sammlungszusammenhang schaffen. Dabei kann allerdings – anders als heute – das ausschlaggebende Kriterium nicht mehr die Lizenzierung von Dokumenten sein,<sup>77</sup> sondern ein inhaltliches Kriterium und ein breiteres Sichten von Titeln unabhängig von Form und Geschäftsmodell sind gefragt.

<sup>77</sup> Häufig werden Open-Access-Dokumente Bibliotheksnutzerinnen und -nutzern, falls überhaupt, lediglich über Discovery-Software (wie beispielsweise Primo) angeboten, jedoch nicht in einer auf das Sammelprofil der Einrichtung zugeschnittenen Auswahl wie lizenzierte Dokumente. Die Tendenz, eine möglichst große Menge statt eine durch Einzeltitelauswahl selektierte Sammlung anzubieten, zeigt sich auch an der Erwerbung von E-Book-Paketen oder dem Einspielen eines großen Titelangebots in den Bibliothekskatalog für die *Patron Driven Acquisition*.