



HAL
open science

Von wissenschaftlichen Editionen als interoperable Projekte, oder: Was können eigentlich digitale Editionen?

Anne Baillot, Markus Schnöpf

► **To cite this version:**

Anne Baillot, Markus Schnöpf. Von wissenschaftlichen Editionen als interoperable Projekte, oder: Was können eigentlich digitale Editionen?. Historische Mitteilungen der Ranke-Gesellschaft, 2015, Die Zukunft der Digital Humanities, Beiheft 90. halshs-01140930

HAL Id: halshs-01140930

<https://shs.hal.science/halshs-01140930>

Submitted on 9 Apr 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Von wissenschaftlichen Editionen als interoperable Projekte, oder: Was können eigentlich digitale Editionen?

Anne Baillot und Markus Schnöpf (Berlin), März 2015

Während gedruckte Editionen eine lange Tradition vorweisen können und über Regelwerk, Referenzen und Normen verfügen, stecken wissenschaftliche digitale Editionen immer noch in den Kinderschuhen.¹ In vielerlei Hinsicht lehnen sich letztere noch an herkömmliche Printeditionen an.² Das Blättern wird imitiert, seltene Unicode-Zeichen werden visuell, aber nicht unbedingt semantisch eingesetzt, um Diakritika und Superskripta über Buchstaben so exakt wie möglich wiederzugeben. Damit wird das digitale Konzept hinsichtlich der Veranschaulichung textueller Elemente an seine technischen Grenzen gebracht, ohne – bei weitem – sein eigentliches Potenzial ausgeschöpft zu haben. Digitale Editionen bieten schließlich ganz andere Möglichkeiten, nicht zuletzt, was die Verbindung eines Abbilds einer Quelle mit dem edierten Text angeht. Diese Möglichkeiten stellen aber einerseits den Bezug zum für wissenschaftliche Printeditionen zentralen „Befund“ (als Gegensatz zur Interpretation) in Frage,³ andererseits verändern sie das Verständnis der Nachnutzung des edierten Textes, der als für sich stehendes, jedoch immer wieder ergänzbares Produkt fungiert.

Digitale Editionen sind in der Regel „Projekte“, d. h. in Zeit und Mitteln beschränkte,⁴ kaum forschungsgetragene, vielmehr als Hilfsmittel geltende Unternehmen, die wenig akademischen Ruhm

1 Auch die deutschsprachige Forschung über digitale Editionen steht noch am Anfang. Die Dissertation von Patrick Sahle, *Digitale Editionsformen zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels*, Norderstedt 2013 gilt nach wie vor als alleiniges Referenzwerk. Von ihm und den Kollegen des Instituts für Dokumentologie und Editorik stammt auch die Definition anhand derer der „Kriterienkatalog zur Besprechung von digitalen Editionen“ (hier Version 1.1.) entwickelt wurde: „Bei einer (wissenschaftlichen) Edition handelt es sich um die Publikation von Informationsressourcen, die eine erschließende Wiedergabe historischer Dokumente und Texte bieten. „Digitale Editionen“ werden nicht nur in digitaler Form publiziert, sondern folgen in ihrer Methodologie einem digitalen Paradigma – so wie traditionelle gedruckte Editionen eine Methodologie verfolgten, die dem Paradigma der Druckkultur entspricht.“ (<http://www.i-d-e.de/publikationen/weitereschriften/kriterien-version-1-1/>). 2007 zweifelte Hans Walter Gabler den Paradigmenwechsel noch stark an: „(...) das Grundverständnis der Editorik von der Hierarchie der Größen ‚Text‘ und ‚Dokument‘ ist bisher noch kaum hinterfragt worden.“ (Hans Walter Gabler, *Das wissenschaftliche Edieren als Funktion der Dokumente*, in: *Jahrbuch für Computerphilologie 8/2006*, online: <http://computerphilologie.digital-humanities.de/jg06/gabler.html>). [Alle online-Ressourcen hier und im Folgenden wurden am 22.02.2015 zuletzt konsultiert]

2 Vgl. Gabler: „Die wissenschaftliche Edition ist bisher noch aus dem materiellen Medium in das elektronische Medium, Element für Element, zumeist analog, übertragen worden.“ (Gabler, *Das wissenschaftliche Edieren*, 61).

3 Die Frage der Interpretation als Teil der editorischen Leistung wird von Bodo Plachta, *Editionswissenschaft. Eine Einführung in Methode und Praxis der Edition neuerer Texte*, Stuttgart 1997, als ein kritisches Debattenfeld erwähnt, ohne dass an dieser Stelle dazu explizit Stellung genommen wird (Plachta, *Editionswissenschaft*, 35-38).

4 Die längerfristigen Editionsprojekte, auch digitaler Art, sind meist an Akademien angesiedelt, etwa die Referenzedition „Carl-Maria-von-Weber-Gesamtausgabe“ (<http://www.weber-gesamtausgabe.de/>). In Bezug auf Druckeditionen schreibt Bodo Plachta: „Editionen gehören, welchen Anspruch sie auch immer haben, zu wissenschaftlichen Großunternehmungen, sind zeit-, arbeits-, personal- und kostenintensiv.“ (Plachta, *Editionswissenschaft*, 11, wobei er ebd. 44 auf die seit den 1970er Jahren



und eine xml-Version der editorischen Arbeit voraussetzen, wobei in letzterer das Markup eine zentrale Rolle spielt.⁹

Dekodieren/Rekodieren: eine neue Sichtbarkeit der Erschließungsprozesse

Nicht alle geisteswissenschaftlichen, textbezogenen Fragestellungen erfordern eine Auseinandersetzung mit der Entstehungsgeschichte eines gedruckten Werkes. Das Werk als abgeschlossenes Ganzes steht bei den meisten textbezogenen Forschungsverfahren im Mittelpunkt, d. h. durch mögliche Varianten eingeführte, jedoch verglichen mit der gesamten Textmenge geringfügige Abweichungen gelten insbesondere im Kontext von quantitativen Textanalysen (einer beliebten Spielweise der textbezogenen Digital Humanities) als irrelevant. Insofern sind Fragen der Textaufbereitung, wie sie im Rahmen von wissenschaftlichen digitalen Editionen im Mittelpunkt stehen, nur für eine kleine Sparte der geisteswissenschaftlichen Forschung von Relevanz. Sie führen jedoch eine gelehrte Tradition der Textbereitstellung fort, die mindestens auf Karl Lachmann und mit ihm auf die Ursprünge der für den deutschsprachigen Raum brisantesten philologischen Tradition zurückgeht¹⁰, und reflektieren diese im neuen medialen Kontext. Die Kollation mehrerer Textzeugen¹¹ wird durch computergestützte Verfahren erleichtert. Digitale Werkzeuge bieten darüber hinaus die Möglichkeit, jeden einzelnen Textzeugen für sich darzustellen und ihm so eigenständiges Gewicht zu verleihen. Somit gewinnt der Rekurs auf ebendiesen (etwa handschriftlichen) Textzeugen eine erneuerte Bedeutung..

Wird dem transkribierten Text ein Bild der dazugehörigen Handschrift gegenübergestellt¹², ergibt sich für den Leser die Möglichkeit, den Prozess der Textkonstitution, und insbesondere seiner

9 Zur Rolle von Markup in wissenschaftlicher Textaufbereitung vgl. Baumann, *Interchange*, insbesondere die Definition von Interoperation im Gegensatz zu negotiated interchange und blind interchange, wobei Baumann unterstreicht, dass XML/TEI nur blind interchange ermöglicht (so auch Peter Stadler als Kommentar dieser Definition in: Peter Stadler, Interoperabilität von digitalen Briefeditionen, in: Hanna Delf von Wolzogen / Rainer Falk (Hg.), *Fontanes Briefe ediert*, Würzburg 2014, 278-287. 2010 beschreibt Susan Schreibman die Text Encoding Initiative als ein „interoperable framework for information exchange“ (Susan Schreibman, The Text Encoding Initiative. An Interchange Format Once Again, in: *Jahrbuch für Computerphilologie* 10/2010, 13).

10 In seiner kritischen Diskussion unterschiedlicher Editionsansätze setzt sich David C. Greetham mit G. Thomas Tanselles Definition des Zieles editorischer Unterfangen auseinander und zitiert dabei Tanselles Stellungnahme in *Textual Criticism and Scholarly Editing*: „[I] meant only to assert that, if one were undertaking a critical edition, the most appropriate goal would seem to be the reconstruction of an authorially intended text (leaving other goals to other kinds of editions).“ David C. Greetham, *Theories of the Text*, New York 1999, 169. Mit der Frage der Autorität setzt sich die editionswissenschaftliche Forschung in den unterschiedlichen Sprachräumen auf der Grundlage unterschiedlicher Ansätze auseinander, findet sich aber immer in der einen oder anderen Form wieder.

11 Als Primäraufgabe des Editors definiert Louis Hay eine solche Kollation von Textzeugen: „(...) le premier travail de l'éditeur consiste toujours à *établir* un texte. Il s'agit d'une part d'identifier les variantes d'origine étrangère (pour les signaler, tout en les éliminant du texte); de l'autre, les variantes d'auteur (pour les signaler et les faire apparaître dans l'édition sous une forme appropriée).“ (Louis Hay, *La Naissance du Texte*, Paris 1989, 53.)

12 Ein herausragendes Beispiel einer solchen Gegenüberstellung zeigt die Edition der Briefe Vincent van Goghs: <http://www.vangoghletters.org/>.



Mittelpunkt und nicht das fertige Produkt Drucktext, geschweige denn das kanonisierte Werk.¹⁸ Allerdings stieß die Realisierung dieses Ansatzes auch an ihre Grenzen: Soll jedem Element aus jedem Textzeugen prinzipiell eine gleichwertige Bedeutung beigemessen werden, dann muss auch jeder Kringel editorisch wiedergegeben werden. Lohnt sich der Aufwand bei nicht-kanonischen Autoren? Bezeichnenderweise stehen im Mittelpunkt zahlreicher, mit den Prinzipien der *critique génétique* arbeitenden editorischen Vorhaben ausgerechnet unumgängliche Größen des Literaturkanons.¹⁹ Wenn auch nur implizit oder auf pragmatische Gründe zurückgeführt (ediert wird das, wozu aussagekräftiges Material zur Verfügung steht), wird dabei der Tradition Rechnung getragen. Sollte man es dann mit Autoren zu tun haben, die ihre Arbeitsunterlagen sorgfältig aufbewahrt haben – so wie es bedeutende Schriftsteller des Öfteren tun –, ist dem Anspruch auf Totalität in der Wiedergabe des Textentstehungsprozesses kaum gerecht zu werden.

In Printeditionen ist die Wiedergabe einer dynamischen Textstruktur, die unterschiedliche Bearbeitungsschichten, -kontexte und -formen berücksichtigen möchte, durch die lineare Form des Buches so leicht nicht wiederzugeben. Dies gilt allerdings nicht nur für hochkomplexe Editionen, die den Ansprüchen der *critique génétique* gerecht werden wollen bzw. mehrere Textzeugen heranziehen: Auch bei nicht-textgenetischen Editionen ist das Blättern zu den am Ende eines Bandes bzw. in einem separaten Band befindlichen Verzeichnissen (Personen, Werke, Orte, Schlagworte) eine mühselige Angelegenheit und das Einarbeiten in die editionseigenen Konventionen zeitaufwendig.

Digitale Editionen bieten eine Grundstruktur, die es möglich macht, eine Reihe von praktischen Schwierigkeiten zu umgehen und damit effizient Informationen zu vermitteln, die mit der Erschließung des Textes im Zusammenhang stehen. Das Klicken potenziert das Blättern, indem eine beliebig tiefe Verschachtelung hinter dem anklickenden Element zur Verfügung gestellt werden kann. Die Leseerfahrung allerdings ändert sich dahingehend, dass ausgerechnet die anklickbaren Elemente farblich formatiert werden und sich mit anderen, aus semantischen Gründen hervorgehobenen Stellen überlagern. Insofern lenkt der Editor in digitalen Editions-umgebungen unvermeidlich eine bestimmte Lektüre des Textes. Außerdem wird von Seiten von Webdesignern oft damit argumentiert, dass online-Lesekomfort nicht zuletzt im Erreichen einer Information durch möglichst wenige Klicks besteht, was der Potenzierung der Verschachtelung Grenzen setzt. Insgesamt wird im visuellen Umgang mit digitalen Umgebungen ein optischer Lese- bzw. Nutzungsrahmen²⁰

18 Vgl. Shillingsburg: „In spite of the fact that in the 1980s editorial circles witnessed a paradigm shift in which the concept of a definitive end product was widely replaced by the concept of process in which multiple texts represent the work, nevertheless, the physical limitations of print editions and the linear reading habits of most readers have continued to force the predominance of clear-reading texts as primary feature of new scholarly editions.“ (Shillingsburg, *Scholarly Editing*, 77)

19 So etwa Proust: <http://www.item.ens.fr/index.php?id=578147> oder Flaubert: <http://www.item.ens.fr/index.php?id=115642> sowie <http://www.dossiers-flaubert.fr/>. Auch die in Almuth Grésillon, *La Mise en oeuvre. Itinéraires génétiques*, Paris 2008, herangezogenen Fallbeispiele bewegen sich weitestgehend im Rahmen des Kanons: Flaubert, Zola, Proust, Supervielle, Ponge.

20 Zu den damit zusammenhängenden Nutzungsreflexen und -konventionen vgl. Jacques Bertin, *Sémiologie graphique. Les diagrammes – Les réseaux – Les cartes*, Paris⁴2013. Die Einleitungssätze dieses 1965 im Erstdruck erschienenen Werkes, das Graphik „de-kodiert“, sprechen für sich: „Le graphique tient ses lettres de noblesse de sa double fonction de mémoire artificielle et d'instrument de recherche. Outil rationnel et efficace lorsque les propriétés de la perception visuelle sont pleinement employées, elle fournit l'un des deux

vorausgesetzt, der mehr als ein materieller Zwang denn als ein Baustein im Datenmodell zu betrachten ist.

Das Datenmodell ist nur sekundär am visuellen Ergebnis orientiert, sofern die Editionsdaten tatsächlich strukturiert sind. Die einen online-Text strukturierende Auszeichnung – in digitalen Editionen mittels XML/TEI²¹ – sowie die Art und Weise, wie die unterschiedlichen Auszeichnungsstrukturen aufeinander verweisen, sind es, die das Datenmodell bestimmen. Letzteres ist in seiner Grundanlage nicht primär optisch konzipiert. Es bildet, genau genommen, die Schnittstelle zwischen dem, was der Editor in der ihm vorliegenden Handschrift sieht und dem, was er seinem Leser/Nutzer an Informationen vermitteln möchte. Es steht somit zwischen zwei Momenten der Interpretation. Diese Schnittstelle hat es in der Geschichte der Edition stets gegeben, und sie war immer gewissermaßen ein Kodieren (im Sinne eines De-kodierens und Re-kodierens)²²: Der Editor liest die Handschrift, bereitet eine Rohtranskription vor, markiert Unsicherheiten, kommentarbedürftige Stellen, Emendationen. Dieser Arbeitsschritt hatte früher nur in den persönlichen Archiven der Editoren bzw. der Editionsprojekten ein Dasein. Er existierte nicht als eigene wissenschaftliche, *einsehbar* Leistung. Dies alles umfasst aber heute die XML-Datei: sie ist das Archiv, das zum Aufbewahren dient, sie ist das, was bleibt, das, wonach man die eigentliche wissenschaftliche Leistung über die Zeiten hinweg einschätzen kann. Das Verfügbarmachen des Quellcodes in digitalen Editionen bildet damit eine Art Tabubruch mit bisherigen Gewohnheiten. Der Quellcode ist Teil der Dokumentation des editorischen Prozesses. In vielen Fällen ist die digitale Verzeichnung schematischer, aber auch mathematisch genauer, können doch beispielsweise Unsicherheiten und Vermutungen in der Lesung in Prozentzahlen angegeben werden.

Insofern verwundert es nicht, dass es Editoren gibt, die die XML-Dateien öffentlich zur Verfügung zu stellen, andere dagegen, die sich weigern, dies zu tun. Die Debatte, wozu die Bereitstellung gut sein soll, trifft einen Nerv in der Auseinandersetzung um den medialen Wandel und wird in anderen Wissenschaften im Bereich der Publikation von Forschungsdaten geführt.²³ Die XML-Dateien nicht freizugeben, entspricht gewissermaßen einem seit langem eingebürgerten wissenschaftlichen Gestus: Mein Hintergrundwissen gehört nur mir, mein wissenschaftlicher Wert liegt in dem, was nur

'languages' du traitement de l'information. L'écran cathodique lui ouvre un avenir illimité." (Bertin, *Sémiologie*, 6).

21 Die Text Encoding Initiative bietet ein Regelwerk zur Verwendung der Auszeichnungssprache XML für digitale Editionen an, welches zum De facto Standard geworden ist: <http://www.tei-c.org/Guidelines/P5/>.

22 Vgl. in diesem Sinne Gabler, *Das wissenschaftliche Edieren*, 57: „Das traditionelle Edieren löst also die Texte von den Textträgern ab und lässt effektiv die Dokumente als Dokumente hinter sich. Doch wie sollte das auch anders gehen? Schließlich ist Edieren ja bloß eine besondere Spielart des Überliefers im Abschreiben, also die Übertragung von einem ‚Dokument 1‘ – der materiellen Unterlage, von der sie abgelöst werden – auf ein ‚Dokument 2‘: die Unterlage, auf der sie neu aufgetragen werden.“ sowie zu Edition als „*Konstruktion von Texten*“: „Denn Texte kritisch zu edieren heißt genau dies: sie zu konstruieren.“ (Gabler, *das wissenschaftliche Edieren*, 60)

23 Vgl. exemplarisch Pampel, Forschungsdaten-Repositorien . Informationsinfrastrukturen für nachnutzbare Forschungsdaten, Potsdam 2012, online: <http://gfzpublic.gfz-potsdam.de/pubman/item/escidoc:245340> und für die Psychologie Günther, Open data und data sharing: Neue Perspektiven im Umgang mit Forschungsdaten in der Psychologie, Saarbrücken 2010, online: <http://nbn-resolving.de/urn:nbn:de:bsz:291-psydok-27143>



ich weiß. Nur die Ergebnisse sollen gezeigt werden, nicht der Entstehungsprozess und die Dokumentation derselben. Wer im Gegensatz dazu seine XML-Dateien zugänglich macht, tut dies meistens, damit sie weiterverwendet werden und damit der Arbeitsprozess nachvollziehbar gemacht werden können, im Sinne eines womöglich illusionären, aufklärerischen Fortschrittsgedankens: An meine Arbeit schließt sich noch Besseres an. Den Quellcode zum Bestandteil der editorischen Leistung zu machen, wird nur zu leicht als ehrgeizig wahrgenommen (als ob der Wissenschaftler das „Unsaubere“ für wichtiger hielte als es in Wahrheit sei). Interessanterweise spiegelt sich in dieser Sicht auf die „Black Box“ des wissenschaftlichen Entstehungsprozesses genau das wieder, was bei der Edition von Schriftstellerwerken in der Auffassung der literarischen Textentstehung bei der Edition von Schriftstellerwerken auf dem Spiel steht. In beiden Fällen gilt: Nicht der Prozess ist lesenswert, sondern das fertige Endergebnis. Damit wird auch vorausgesetzt, dass ein fertiges Ergebnis das Maß aller Dinge, ja überhaupt erreichbar sei – die Gegenillusion zum aufklärerischen Fortschrittsgedanken.

Darüber hinaus gibt es bei Weitem nicht nur *eine* Art und Weise, ein bestimmtes Textphänomen digital anzugehen und auszuzeichnen. Es gibt zwar nicht beliebig viele Wege, aber in der Regel – zumal bei komplexeren Textverhältnissen – mehrere.²⁴ Indem man seinen Quellcode zur Verfügung stellt, macht man seine Arbeitsschritte transparent, sich aber auch angreifbar. Einige der dort einsehbaren Auszeichnungsentscheidungen können angezweifelt werden oder kontrovers erscheinen.

Die Entscheidung für eine bestimmte Auszeichnung hängt in der Regel nicht primär vom Einzelfall ab, sondern von der Gesamtstruktur des Korpus und von den Grundsätzen der jeweiligen Texterschließung. Diese führen zu einer Interpretation des zur Verfügung stehenden Auszeichnungsapparates (d. h. der Gesamtsumme der TEI-Regeln). Was man als Editor daraus im eigenen Quellcode verwendet und was der Leser dann in ebendiesem Code sieht, vermittelt ein Textverständnis, es spiegelt die interpretatorische Leistung wieder und fordert zur hermeneutischen Auseinandersetzung auf – in erster Linie auf der Ebene der Auszeichnung.²⁵

Insofern spielt in der Einschätzung der Herangehensweise an den Text neben den XML-Dateien die Dokumentation der Kodierungsentscheidungen eine entscheidende Rolle. Allerdings ist es bislang so, dass diejenigen, denen Texthermeneutik am Herzen liegt, nur in den seltensten Fällen darin geschult sind, XML/TEI-Dokumentationen entschlüsseln zu können. Und gerade, weil digitale Editionsprojekte Projektcharakter besitzen, müsste auf diese Dokumentationen ein besonders starker Akzent gelegt werden, denn die genaue Erfassung der Auszeichnungsentscheidungen ist eine

24 Vgl. hierzu mit Blick auf die Nachnutzungsmöglichkeiten Desmond Schmidt, *Towards an Interoperable Digital Scholarly Edition*, in: *Journal of the Text Encoding Initiative [Online]*, 7/2014. URL : <http://jtei.revues.org/979>; DOI : 10.4000/jtei.979.

25 Es sei hier auf die Diskussionen der unterschiedlichen TEI-Listen verwiesen, die täglich solche Beispiele zutage fördern (die Hauptliste ist zu finden unter: <https://listserv.brown.edu/archives/cgi-bin/wa?A0=TEI-L>).

der wenigen Schutzmaßnahmen, die es ermöglichen, Langfristigkeit überhaupt in den Blick zu nehmen – und dabei insbesondere Langfristigkeit der geleisteten interpretatorischen Textarbeit.²⁶

Die Idee, es komme bei der XML/TEI-Codierung nicht so sehr auf das optische Endergebnis an, mag im Kontext editorischer Arbeit etwas kontradiktorisch klingen. Wer würde sich noch die Mühe machen, wenn am Ende kein Anzeigeversprechen stehen würde? Ist es denn nicht einer der wichtigsten Vorteile digitaler Editionen, dass sie bestimmte Informationen anzuzeigen erlauben – an erster Stelle Verbindungen? Die Diskrepanz zwischen dem Kodierungsakt und der Visualisierung des Kodierten, der Wandlungsprozess zwischen dem Kodierten und dem Angezeigten folgt jedoch keinem linearen zwangsläufigen Translationskonzept, nach dem aus einer bestimmten Kodierungsformel eine bestimmte Anzeige generiert werden würde. Genauso wie unterschiedliche Kodierungen möglich sind, um bestimmte Textphänomene auszuzeichnen, stehen unterschiedliche Visualisierungsoptionen zur Verfügung, um das Kodierte umzusetzen. An dieser Stelle schaltet sich ebenfalls Interpretation dazwischen. Auch dem, der die Online-Visualisierung ausarbeitet – dem Designer – schweben Umsetzungen vor, die vom XML/TEI-Code unabhängig sind. Wir haben es mit mehreren Niveaus von Interpretation zu tun, ehe der Text den edierten Zustand erreicht: Der Editor erschließt und kodiert die Handschrift, der Entwickler/Visualisierer interpretiert das Kodierte und setzt es in eine Visualisierung um. Mindestens zwei Arbeitsschritte und drei mediale Wandlungen sind erforderlich.

Digitale Editionen wandeln den Text in ein Informationsnetz um. Es ist notwendig, dieses zu benennen, denn es sind für jeden Text jeweils unterschiedliche solcher Informationsnetze denkbar. Einige setzen den Akzent auf die Textgenese, andere auf die Entitäten, andere auf linguistische Phänomene diverser Art. Nicht zuletzt stellt sich die Frage, wo diese hermeneutische Selbstpositionierung im Kontext einer digitalen Edition ihren Platz finden kann. Es geht an dieser Stelle dabei nicht mehr nur um die Kodierungsrichtlinien, auch nicht um die Designentscheidungen, und schon gar nicht um die Alternative zwischen diplomatischer Umschrift und Lesefassung. Es geht um die Ausformulierung des Fragenkomplexes, der zwangsläufig mit einer XML-Auszeichnung zusammenhängt. Hieraus ergeben sich allerdings neue Schwierigkeiten: Nicht nur hat eine solche Selbstverortung keinen ihr gewidmeten Ort im editorischen Zusammenhang, sondern ist darüber hinaus die Gefahr kaum zu vermeiden, ausgerechnet die hermeneutische Tiefe der Auszeichnung in einer solchen Artikulation zu verflachen. Die Distanz zum eigenen Textverständnis soweit auszuformulieren, dass sie schon bei ihrem Entstehen ihre eigene Historizität reflektieren kann, ist keine leichte Aufgabe.²⁷

26 In Ermangelung expliziter Standards für die Verfassung solcher Dokumentationen sei an dieser Stelle beispielhaft auf die Kodierungsrichtlinien der Edition *Briefe und Texte aus dem intellektuellen Berlin um 1800* verwiesen, die unter folgender URL im pdf-Format heruntergeladen werden kann: <http://tei.ibi.hu-berlin.de/berliner-intellektuelle/about?de>.

27 Damit hat sich die angelsächsische Editionswissenschaft bereits intensiv beschäftigt; vgl. beispielsweise Greetham: „And thus, while I maintain that theory, a theory, does lie behind every textual operation, from selecting copy-text to charting variants to writing annotations, and that theory is therefore ubiquitous, I also believe ([...]) that theory also has local and specific roots, and that a particular theoretical cast is contingent upon a social and cultural milieu that makes *this* theory at *this* time utterable and plausible.“ (Greetham, *Theories*, 20) und „This critical bind obviously becomes the most consistently problematic issue in textual criticism as the historical distance between the preserved object of study (the



Textauszeichnung trägt immer das Gepräge eines Textverständnisses, führt dieses Verständnis in den Vordergrund. Wenn XML-Dateien frei zugänglich und wieder verwendbar sind, hat jeder die Möglichkeit, daraus das zu machen, was er möchte: das vorhandene Markup zu verwerfen, sein eigenes einzubauen, seine Edition anzubieten.²⁸ Diese Vorstellung führt allerdings zu einer merkwürdigen Aporie. Denn wenn man davon ausgeht, dass ein Text sich aus einer von der Auszeichnung bereinigten XML-Datei extrahieren lässt, folgt man genau dem, was der Eklektizismus immer gemacht hat, nämlich von der Idee geleitet zu werden, dass es letztlich *einen* Text gibt. So bleibt zu hoffen, dass die Verbreitung von CC-BY-Lizenzen nicht zu einem Verflachen der hermeneutischen Leistung von Text-Markup führen, sondern, im Gegenteil, der Interpretationsfülle, den vielen Sichtweisen ein Gesicht geben wird.²⁹

Das Kodierte zeigen

Digitale Interfaces bieten zahlreiche Möglichkeiten, sich von einer bloßen Buch-Imitation zu lösen. Sicherlich sind diese Möglichkeiten eingeschränkt, wenn man sich an der Handschrift orientiert, denn das Blatt, mit seiner Dimension in Länge und Breite, ist in der Regel ebenso der materielle Träger der Handschrift, wie es der Träger für Bücher ist.³⁰ Eine Edition, die zwar einen Text zur Verfügung stellt, aber nicht nah an der Originalhandschrift arbeitet, kann sich von dem vorgegebenen Format der Seite lösen. Jedoch ergeben sich daraus nicht unerhebliche Fragen, was beispielsweise die Zitierbarkeit einzelner Textstellen angeht. Wie zitiert man aus einer ePub, die für den Kurzsichtigen den Text auf dem Träger anders verteilt als für den Weitsichtigen?³¹

Grundsätzlich bietet die dynamische Darstellung von Textinhalten die Möglichkeit, eine Überfrachtung der Textansicht zu vermeiden, indem eine Reihe von Informationen nur als anklickbare Optionen zur Verfügung gestellt werden, sei es in der bereits erwähnten Form der verschachtelten Klicks oder in der Gestalt von fakultativ anzeigbaren Zusatzinformationen. Bei der Implementierung dieses digitalen Handwerkszeugs stehen editorische Grundregeln und Designprinzipien oft im Gegensatz zu einander. Pop-Ups haben beispielsweise den – aus editorischer Sicht kaum nachvollziehbaren – Nachteil, dass sie in dem Moment, wo sie erscheinen, den Text unter sich verbergen. Ebenso verschieben alle farblichen Hervorhebungen (wie sie beispielsweise zur Identifizierung von Links nötig sind) die Aufmerksamkeit auf eine Art und Weise, die nicht

manuscript) and the inferred moment of composition grows ever larger; and it is no surprise that hermeneutics as either a technical or an aesthetic manifestation of intentionalism should occur most prominently in those disciplines in which the gap was largest – biblical and classical studies – thereby reinforcing the cultural, moral and philosophical rationale for their study.“ (Greetham, *Theories*, 175).

28 Dagegen wendet allerdings Peter Stadler ein, dass Markup zu stark in den Text eingebettet sei, um von diesem getrennt zu werden (Stadler, *Interoperabilität*, 287).

29 Die unterschiedlichen CC-BY-Lizenzen und ihre Bedeutungen sind zu finden unter: <http://creativecommons.org/>.

30 Nach einer umfangreichen Schilderung der Rolle der Seite von der Antike bis ins moderne Zeitalter (93ff.) schlussfolgert Louis Hay: „(...) on peut supposer que l'informatique induira, chez un homme que l'on pourrait qualifier d' „écranique“, un changement profond de cet habitus, une conception de l'écrit qui risque de ne tenir en rien à notre idée du textuel. C'est au sein de l'écrit, de la trace et de l'inscription que tout se joue. Et qu'une phase historique s'achève (...).“ (Hay, *La Naissance*, 108).

31 Zu den im Allgemeinen mit dem Zitieren digitaler Dokumente zusammenhängenden Fragen vgl, s. Klaus Prätor, Zur Zukunft des Zitierens. Identität, Referenz und Granularität digitaler Dokumente, in: *editio* 25/2011, 170 – 183 (DOI 10.1515/edit.2011.010).

textinhärent ist. Überhaupt kann die Verwendung von Farben – digitale Editionen sind in der Regel nicht schwarz-weiß, während diese Farbkombination bei Printeditionen die Regel (mit Ausnahmen) ausmacht – durchaus als interpretatorischer Eingriff betrachtet werden.³² Mit der Entscheidung für eine Farbpalette wird eine ganze Reihe von Konnotationen aufgefächert, die weder kulturell noch browsertechnisch universell einsetzbar sind. Nicht nur schafft eine Vervielfältigung der farbigen Markierungen Unruhe auf einer Webseite, sie vermittelt unter Umständen zudem nicht die gewünschte Botschaft. Auch dies relativiert die Orientierung am editionswissenschaftlich maßgeblichen Befund.

Design besteht nicht nur und nicht primär aus Text, was für den Zusammenhang digitaler Editionen mit dem Webdesign Fluch und Segen zugleich ist. Das Verhältnis von Text und Bild verschiebt sich so, dass im Rahmen der editorischen Leistung dem Bild mehr zugetraut wird als es in Büchern in der Regel der Fall ist. Wordclouds lesen sich nicht linear, Netzwerkvisualisierungen, Karten und Zeitleisten auch nicht, dabei sind sie es, die in digitalen Editionen nicht selten für die Gesamtorientierung sorgen.³³

Maschinen lesen

Das menschliche Leben ist begrenzt, somit auch die Menge an Büchern, die in einem Menschenleben gelesen werden können. Mit der Digitalisierung wird die Menge an Büchern, die der Mensch lesen kann, zwar nicht größer, aber durch die maschinellen Lesemöglichkeiten des Computers ergänzt. Gregory Crane stellte sich schon 2006 der Frage, wie sich das Leseverhalten angesichts der fortschreitenden Digitalisierung unserer Bibliotheken ändert.³⁴ Vor allem die Anstrengungen, die im US-amerikanischen Raum unternommen werden, um komplette Bibliotheksbestände zu retrodigitalisieren – hier ist vor allem Google Books zu nennen – erhöhen den Zugang zu Wissensbeständen, die in der analogen Welt unvorstellbar gewesen wären. Auf europäischer Ebene spielt als Portal für verschiedene Datenlieferanten die virtuelle Bibliothek Europeana eine wichtige Rolle.³⁵ So erfreulich es ist, Literatur nun nur noch einen Klick entfernt auf den Bildschirm rufen zu können, so werden dabei doch einige Fragen aufgeworfen, die das Leseverhalten betreffen. Google Books erfüllt derzeit noch nicht die Standards, die an eine Ressource für die Wissenschaft gestellt werden. Vor allem zwei Gründe sprechen dagegen: Einerseits wird der Volltext durch OCR erstellt. Die Verfahren werden zwar immer genauer, aber auch eine Fehlerquote von zwei Prozent ist für eine genaue wissenschaftliche Analyse zu hoch.³⁶ Zudem führt die automatisierte Texterfassung gerade bei älteren Drucken zu einer höheren Fehlerquote. Der zweite Grund liegt bei Google selbst. Startete Google Books ursprünglich als offenes Projekt, das zwar von einer nicht in öffentlicher Hand befindlichen Firma finanziert wurde, so sind inzwischen aus Gründen der Marktbildung inzwischen

32 Ähnliches vermerkt allerdings Plachta in Bezug auf Druckeditionen frühneuzeitlicher Texte (Plachta, *Editionswissenschaft*, 21–22).

33 Vgl. die Startseite der August Wilhelm Schlegel-Briefedition (<http://august-wilhelm-schlegel.de/briefedigital/>) sowie die Urkunden Kaiser Karls IV. mit einem zeitlich-kartografischen Zugang (<http://telota.bbaw.de/constitutiones/#timeMap>).

34 Gregory Crane, What do you do with a Million Books, in: *D-Lib Magazine* 12/2006, online: <http://www.dlib.org/dlib/march06/crane/03crane.html>.

35 <http://www.europeana.eu>.

36 Bei einer durchschnittlichen Zeichenzahl von 4000 Charakteren pro Seite sind das 80 Fehler pro Seite.



viele Volltexte nur noch Klappentexte und werben für den Kauf des betreffenden Werkes.³⁷ Das Angebot selbst kann jederzeit wieder eingestellt werden, wenn sich der Betrieb für Google selbst nicht mehr lohnen sollte. Dennoch hat Google Books auch quantifizierende Methoden in die Literaturforschung populär werden lassen. Anhand ngrams können zeitliche, somit historische und kulturelle Tendenzen beliebiger Begriffe im Gesamtkorpus der digitalisierten Werke der letzten zweihundert Jahre recherchiert werden.³⁸ Die Bücher selbst werden dabei nicht mehr linear gelesen, sondern anhand computerlinguistischer Methoden massenausgewertet.

Die Digital Humanities gehen mit der Einführung quantifizierender Methoden in den Geisteswissenschaften einher. Für die Geschichtswissenschaften und die Computerlinguistik ist dies beispielsweise nichts Neues, wurden doch im Rahmen der Sozialgeschichtsschreibung vor allem in den 1980ern im Rahmen der historischen Demografie Reihenauswertungen von Kirchenbüchern vorgenommen. Die historische Demografie hat auch gezeigt, dass sich quantifizierende und hermeneutische Methoden in den Geisteswissenschaften nicht ausschließen.³⁹

In der amerikanischen Literaturkritik hat vor allem Franco Moretti's „distant reading“ zu einer Methodenerweiterung geführt.⁴⁰ Auch auf der 2012 stattgefundenen Tagung der Modern Language Association wurde die Hinwendung der Literaturforschung zu den Digital Humanities manifest:

„So what exactly is that new insurgency? What rough beast has slouched into the neighborhood threatening to upset everyone's applecart? The program's statistics deliver a clear answer. Upward of 40 sessions are devoted to what is called the “digital humanities,” an umbrella term for new and fast-moving developments across a range of topics: the organization and administration of libraries, the rethinking of peer review, the study of social networks, the expansion of digital archives, the refining of search engines, the production of scholarly editions, the restructuring of undergraduate instruction, the transformation of scholarly publishing, the re-conception of the doctoral dissertation, the teaching of foreign languages, the proliferation of online journals, the redefinition of what it means to be a text, the changing face of tenure — in short, everything.“⁴¹

Moretti's Buch behandelt vor allem literarische Kanonbildung. Mit Rückgriffen auf die Evolutionstheorie geht Moretti darauf ein, warum nur ein Prozent der publizierten Belletristik sich langfristig in den Kanon erhaltener Literatur einbringen konnte. Mittels Reihenauswertung von 7000 Titellängen kann Moretti darstellen, dass die Wortanzahl der Titel im Laufe des 19. Jahrhunderts auf

3737 Als Beispiel mögen hier die eigentlich gemeinfreien Anmerkungen von Johannes Bolte und Georg Polivka zu den Kinder- und Hausmärchen der Brüder Grimm dienen, ein Reprint eines Werkes, welches der public domain unterliegt, aber nur in Auszügen angezeigt wird.

https://books.google.de/books?id=Cm_Op7Ur7UC&lpg=PP1&dq=Bolte&hl=de&pg=PP1#v=onepage&q=Bolte&f=false

38 <https://books.google.com/ngrams>.

39 Vgl. exemplarisch: Imhof, *Einführung in die historische Demographie*, München 1977 und Imhof, „*Die Kunst des Sterbens. Wie unsere Vorfahren sterben lernten*“, Stuttgart 1998. Zu bemerken ist allerdings, dass die historische Demografie in Deutschland kaum noch gelehrt wird.

40 Moretti, *Distant Reading*, London 2013. Auf den Webseiten des von Moretti geführten Stanford Literary Lab wird die Diskussion in Form von Pamphleten weiter geführt. Vgl.

http://litlab.stanford.edu/?page_id=255.

41 Stanley Fish, *The Old Order Changeth*, in: New York Times vom 26.12.2011, online:

<http://opinionator.blogs.nytimes.com/2011/12/26/the-old-order-changeth/>.

kurze Titel reduziert wurde, um diese einprägsamer zu gestalten und somit deren Verkäuflichkeit zu verbessern.⁴² Ohne analytische Rechentechnik wäre diese Auswertung nur mit erheblicher Zeitinvestition möglich gewesen. Ein zweiter Punkt betrifft die historische Netzwerkanalyse, die Moretti an den Charakteren von Shakespeares *Hamlet* anwendet. Visualisierungstechniken lassen sich hier für neuartige Analysezwecke verwenden, um die Verflechtungen und Gruppierungen der Akteure des Werkes fassbar zu machen.

Interoperabilität als exemplarische Realisierung der nichtlinearen Anlage digitaler Editionen

Im Gegensatz zu Printeditionen sind digitale Editionen dadurch charakterisiert, dass sie dynamisch angelegt sind. Nicht selten ist es der Anspruch einer Referenzausgabe in Print, endgültigen Charakter zu besitzen. Digitale Editionen ihrerseits bestehen nicht darauf, dass sie alles in sich enthalten, sondern sie sind vielmehr zu anderen Ressourcen hin geöffnet.⁴³ Ihre Sichtbarkeit bzw. Effizienz misst sich nicht zuletzt an ihrer Fähigkeit, Anschlüsse zu finden, zu strukturieren und zu implementieren. Dies bedeutet einmal, Struktur und Qualität der vorhandenen Ressourcen zu kennen, aber auch in einem Akt der Spekulation imstande zu sein, im Datenmodell Raum für den Anschluss an Ressourcen zu lassen, die in der Zukunft liegen mögen – ob sich diese Spekulationen bewahrheiten oder auch nicht.⁴⁴ Insofern besteht diese Arbeit darin, Mosaiksteine so zu legen, dass sich ein Bild formt, welches potenziell erweitert werden kann und bei jeder Ergänzung eine sinnvolle Gestalt annimmt. Diese Struktur ist per se nicht linear, sondern basiert auf der inhärenten Annehmbarkeit von Leerstellen. So paradox es auch beim ersten Blick scheinen mag, sind Nicht-Linearität und Interoperabilität eng mit einander verwoben, wenn es um die Struktur einer wissenschaftlichen digitalen Edition geht. Verlinkungen und die durch Formatähnlichkeit erreichbare Ressourcenzusammenführung spielen dabei eine zentrale Rolle.

Verlinkungen können entweder von einer digitalen Ressource her oder auf eine digitale Ressource hin erfolgen. Sie ermöglichen es, Informationen in dichter Form zusammenzuführen und lange Ausführungen zu vermeiden, indem auf eine primäre, sekundäre oder tertiäre Informationsquelle verwiesen wird. In digitalen Briefeditionen kann beispielsweise bei Briefen auf biographische Informationen hingewiesen werden, bei der Edition von Schriften auf die Wiederholung von Motiven oder aber können korrespondierende Textstellen mit einander verlinkt werden. Der reine Verweis

42 Moretti, *Reading*, 179-210.

43 Mindestens die Anbindung an Bibliotheksstrukturen gehört zu den von der Deutschen Forschungsgemeinschaft dargelegten Evaluationskriterien für wissenschaftliche digitale Editionen (http://www.dfg.de/foerderung/grundlagen_rahmenbedingungen/informationen_fachwissenschaften/geistes_sozialwissenschaften/kriterien_begutachtung_wissenschaftlicher_editionen/index.html).

44 So Bauman auch: „But I hold that interoperationality is not just about use of XML data in current predicted applications (I will want to use this file in an SVG viewer), but rather is also concerned with the use of data with both unintended applications (e. g., linguistic analysis of survey questions) and future applications (e. g., submitting a set of vector graphics images to software that makes a mosaic of them). Notice that I am being (deliberately) self-contradictory, here. I am suggesting that interoperability is always contextual, that we need to know what application we are trying to operate with in order to measure our success at interoperationality; and simultaneously that in order to be considered interoperational, a document should work with applications that are not only unforeseen, but haven't even been written yet. If the application has not been written yet, it is obviously impossible to know whether or not our data will work with it. It is this very contradiction that drives me to say that interoperationality is hard. We cannot know the various contexts in which we would like our files achieve interoperationality.” (Bauman, *Interchange*)



auf andere digitale Ressourcen stellt niedrige technische Anforderungen. Es ist vergleichsweise unproblematisch ist, eine URL in einen xml- oder einen html-Code einzubauen.

Ehe ein solcher Schritt vorgenommen werden kann, ist jedoch eine Reihe von Qualitätssicherungsmaßnahmen nötig. Erstens gilt, Links zu vermeiden, die nach kurzer Zeit mangels Webseitenpflege ins Nichts führen. Die Stabilität einer online-Ressource einzuschätzen, ist nicht immer leicht. Grundsätzlich sollte prinzipiell nur mit persistenten Identifikatoren wie Digital Object Identifiers (sogenannte DOIs) gearbeitet werden, denn bei URLs muss der Anbieter durch Umleitungen im Webserver dafür sorgen, dass sie zitierfähig bleiben. Die momentan geübte Praxis, eine Webseite in Verbindung mit einem Sichtungsdatum in einem Drucktext als Quelle heranzuziehen, kann nur als temporäre Zwischenlösung gelten, wenn sie auch im universitären Bereich am weitesten verbreitet ist. Einen Lösungsansatz, der aber nicht im wissenschaftlichen Kontext erarbeitet worden ist, bietet sicherlich die Verlinkung des bei archive.org gespeicherten zeitlichen Schnappschusses einer Webseite (Wayback Machine) – mit dem Nachteil, dass es sich bei diesem System nicht um eine Initiative der öffentlichen Hand handelt. Durch die lange Präsenz von archive.org (seit 1996) und die seit Kurzem in Angriff genommene Spiegelung der Daten in die Bibliothek von Alexandria ist jedoch kaum davon auszugehen, dass dieser Dienst nur kurzfristig zur Verfügung stehen wird.⁴⁵

Auch im analogen Bereich wird zitiert und verwiesen. Diese Technik wurde ins Digitale übernommen, mit dem Nachteil, dass schon die analoge Zitation in der Regel nur ungenau angewendet wurde. Im Allgemeinen wird dabei auf eine Druckseite referiert, der Rezipient muss also in den durchschnittlich 4000 Zeichen einer Seite die betreffende Stelle suchen, um die Referenz nachlesen zu können. Bis vor Kurzem wurde in der digitalen Übertragung nur auf ein gesamtes Werk in PDF-Form mittels URN⁴⁶ verwiesen. Mit URN granular wurde dieser unbefriedigende Zustand verbessert. Ermöglicht wird nun die Referenzierung auf einzelne Seiten, womit zumindest die Verweisteknik des Printbereichs im digitalen Medium repliziert wird.⁴⁷ Eine weitere Entwicklungsstufe steht allerdings noch an, nämlich die wortgenaue Referenzierung, die bereits in der Alten Geschichte angewendet wird. Dieser Dienst wurde im Umfeld der Homer-Studien entwickelt.⁴⁸ Auch für stark kanonisierte Texte (wie heilige

45 Archive.org als gemeinnützige Organisation möchte laut Eigenaussage die historischen Zustände des Internets als kulturelle Errungenschaft für zukünftige Generationen erhalten sowie einen universellen Zugang zu menschlichem Wissen schaffen. Vgl. http://archive.org/about/bibalex_p_r.php Solange die Nationalbibliotheken noch keine zuverlässigen Prozesse zur Sicherung der jeweiligen Webseiten erarbeitet haben, ist das Angebot von archive.org alternativlos.

46 Uniform Resource Name, ein Auflösungsdienst der Deutschen Nationalbibliothek, der den Namen mit einer Adresse im Internet verbindet (http://www.dnb.de/DE/Standardisierung/PI/pi_node.html). Die meisten elektronischen Dokumentenserver im deutschsprachigen Raum verwenden URNs, um ihre Inhalte langfristig referenzierbar zu gestalten.

47 Vgl. Dorothea Sommer, Christa Schöning-Walter, Kay Heiligenhaus, URN Granular: Persistente Identifizierung und Adressierung von Einzelseiten digitalisierter Drucke, in: *ABI-Technik* 28/2011, 106–114.

48 Vgl. hierzu <http://www.homermultitext.org/hmt-doc/cite/texts/cts.html> als Teil der CITE-Architektur. D. Neel Smith, Gabreil A. Weaver, Applying Domain Knowledge from Structured Citation Formats to Text and Data Mining: Examples Using the CITE Architecture, in: *Dartmouth Computer Science Technical Reports TR2009-649*, online: <http://katahdin.cs.dartmouth.edu/reports/TR2009-649.pdf>.

Schriften) erhöhen solche genauen Referenzierungsmodelle die Vernetzungsmöglichkeiten digitaler Publikationen in qualitativ hochwertige Wissenssysteme.

Dieser Fragenkomplex betrifft ebenfalls die Zitierfähigkeit von Tertiärquellen. Wikipedia verfügt, wenn nicht über die wissenschaftliche Ausgewiesenheit, doch über strukturelle Merkmale für Qualität: Sie ist auf Nachvollziehbarkeit ausgerichtet, sowohl was Informationsquellen als auch was Ausarbeitungsabläufe angeht – was mit Blick auf digitale Zuverlässigkeit ein ungemeiner Vorteil ist.⁴⁹ Dies hat nicht zuletzt damit zu tun, dass die Wikipedia grundsätzlich von ihrer eigenen Perfektibilität ausgeht. Ebenso geht Wissensverlinkung von einer von den Geisteswissenschaften in der Regel wenig gepflegten Prämisse aus, nämlich, dass die Gemeinschaft mehr weiß als das Individuum – eine Prämisse, die so irritierend wirkt, dass oft übersehen wird, dass Autorschaftzuordnung zu den Grundregeln von Wikipedia gehört. Nicht-Linearität heißt auch, dass nicht *ein* Wissenschaftler *einen* Punkt ausarbeitet, sondern dass die Kompetenz verteilt und dort angesetzt wird, wo sich Wissen befindet, anstatt von einer einzigen Feder summiert und wiederholt zu werden. Dieses lässt sich jedoch nur durch die Überwindung des zwischen den Wissensquellen bestehenden Hiatus' bewerkstelligen.

Solche Ansprüche machen eine der größten Anforderungen an digitale Editionen aus: Wie können sie gewährleisten, dass sie referenzierbar sind und bleiben? Ein Teil des Problems liegt in den Fördermöglichkeiten, die zeitlich begrenzt sind und den Projektcharakter von digitalen Editionen festschreiben, aber auch in den institutionellen Bedingungen, die weder zuverlässige Repositorien anbieten, noch Langzeitverfügbarkeit sichern können. Ein weiteres Hindernis stellt die Lizenzierungspraxis dar. Wer seine Inhalte nicht unter einer einfachen CC-BY-Lizenz veröffentlicht, sondern ein –SA bzw. ein –NC hinzufügt (wenn nicht gar beides), macht die Bedingungen ihrer Weiterverwendbarkeit so komplex, dass diese dadurch gefährdet wird. Interoperabilität heißt grundsätzlich: offene Weiterverwendungsbedingungen mit der Garantie einer nachvollziehbaren Referenzierbarkeit.

Um große Textmengen so aussagekräftig zu machen, dass sie imstande sind, geisteswissenschaftliche Fragestellungen zu beantworten, müssen viele Ressourcen möglichst komplex aneinander angeschlossen werden können. Die Maschinenlesbarkeit unterschiedlicher, zahlreicher Quellen kann nur durch die Einhaltung gemeinsamer Standards und Normen angegangen werden, auch wenn diese nicht versprechen, alle Herausforderungen plötzlich zu überwinden. Solche Standards sind zum einen die ISO-Normen⁵⁰, die im Bereich der Textaufbereitung bereits Erhebliches geleistet haben. Es sind aber zum anderen auch Formate wie die TEI-Syntax für XML oder Schnittstellen wie sie Normdateien zur Verfügung stellen.⁵¹ Ein Grundgefüge steht damit für die Realisierung von Interaktionen von

49 Zur Struktur von Wikipedia in einer wohl wenig kritischen, dennoch informativen Selbstpräsentation, vgl. Patrick Danowski und Jakob Voß, Das Wissen der Welt – Die Wikipedia, in: *Open Source Jahrbuch 2005*, 393 – 405.

50 Die Bandbreite der Bereiche, in denen die ISO-Gemeinschaft tätig ist, lässt sich ihrer Webseite entnehmen: <http://www.iso.org/iso/home.html>. Für den geisteswissenschaftlichen Bereich sind ISO-Normen wie CIDOC-CRM als konzeptionelles Referenzmodell für Konzepte und Relationen in der Dokumentation kultureller Überlieferungen, vornehmlich in der Archäologie, relevant. (Vgl. <http://www.cidoc-crm.org>)

51 Vgl. Peter Stadler, Normdateien in der Edition, in: *editio 26/2012*, 174–183.



digitalen Ressourcen untereinander zur Verfügung, welches allerdings noch nicht durchgängig standardisiert ist. Es ermöglicht die eigentliche Potenzierung digitaler Ressourcen und insbesondere Editionen – diesmal nicht nach unten, in immer tiefer verschachtelte Klicks, sondern nach oben, in sich immer weiter verzweigenden Informationsnetze.

Man nehme Egodokumente jeder Art:⁵² Ohne biographisch gesichertes Hintergrundwissen sind diese für Forschungszwecke kaum brauchbar. Der erste Schritt zur Versammlung biographischer Informationen ist die sichere Identifikation einer Person samt ihrer Lebensdaten. Allein das bedeutet in der Regel einen vergleichsweise großen Aufwand. Die einmalige Identifizierung und Referenzierung von historischen Akteuren ist eine Dienstleistung, die in Deutschland in die Gemeinsame Normdatei (GND) einfließt. Diese ermöglicht die Zusammenführung einer beeindruckenden Menge an Informationen.⁵³ Schwierig bleibt, die Ergebnisse der Forschung in diese bibliothekarisch konzipierte und entwickelte Struktur aufzunehmen. Eintragsberechtigt waren in der Vergangenheit in Deutschland nur Bibliotheken; seit ein paar Jahren wurde aber auf Kritik seitens der Forschung reagiert und die GND für weitere Datenbeiträge geöffnet. Dies führt nun zu zahlreichen offensichtlichen Doubletten, die die eindeutige Identifikation der einzeln erfassten historischen Persönlichkeiten erschweren. Die redaktionelle Prüfung der einzelnen Datensätze, die im bibliothekarischen Bereich stattfindet, kann nur mit Blick auf den Quellcode der Daten erkannt werden. Eine maschinelle Lesung ist hiervon unbenommen. Moderne NER-Systeme⁵⁴ sind bei unterschiedlicher Fehlertoleranz in der Lage, Personen in Texten automatisch zu identifizieren, die Hermeneutik verschiebt sich in die Überprüfung der automatisch generierten Ergebnisse. Weitere Verzerrungen gehen mit dem bibliothekarischen Schwerpunkt einher: Erfasst wurden primär Personen, die als Autoren von Werken in den bibliothekarischen Beständen verzeichnet sind.⁵⁵ Insofern wird in dieser Form weiter dem Kanon das Wort geredet, ohne dass ein historisches Korrektiv realistischerweise angegangen werden könne.

Selbst ohne das womöglich moralisch weder vertretbare noch wünschenswerte Ziel zu verfolgen, die Historiographie eines Besseren zu belehren, wäre es jedoch denkbar, dass Bibliotheken, Archive und Forschung darauf hinarbeiten, digital etwas anderes anzubieten als sie es jahrhundertlang im Printmedium getan haben, um so neue Formen der Zugänglichkeit und Interpretation zu schaffen. Interoperabilität⁵⁶ als Kardinaltugend digitaler Editionen wäre damit nichts anderes als der strukturierte und reflektierte Umgang mit der eigenen Verbesserbarkeit und Ergänzenbarkeit sowie mit einem Leserhorizont, der den eigenen Fachbereich übersteigt.

52 Zum Dokumententyp vgl. Stadler, *Normdateien*, 176 mit besonderer Berücksichtigung des Quellentyps in Bezug auf Normdateien.

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

54 Named Entity Recognition-Systeme erleichtern die Identifikation einzelner Personen, wie z. B. Stanfords Named Entity Recognizer (<http://nlp.stanford.edu/software/CRF-NER.shtml>).

55 Erst in jüngeren Jahren wurde das Ziel in Angriff genommen, diesen Schwerpunkt auszuweiten, ebenso wie die Öffnung der nationalen Sammlungsinteressen zu internationalen, interoperablen Formaten wie RDA-MARC 21 den stetigen Vernetzungscharakter einzelner Wissensbasen dokumentiert.

56 Oder „blind exchange“, vgl. Bauman, *Interchange*.