

“Eine digitale Edition kann man nicht sehen” - Gedanken zu Struktur und Persistenz digitaler Editionen

Staecker, Thomas

thomas.staecker@ulb.tu-darmstadt.de
ULB Darmstadt, Deutschland

Einleitung

Digitale Editionen haben nach einer Phase des Ausprobierens und Entwickelns nunmehr eine Reife erreicht, dass sie in vielen Disziplinen nicht mehr als exotischer Sonder-, sondern als Regelfall angesehen werden, was sich in Publikationen wie (Apollon et. al 2014; Driscoll/Pierazzo 2016) oder Förderbedingungen (DFG 2015) spiegelt. Trotz dieser greifbaren Fortschritte stehen digitale Editionen nach wie vor in der Kritik. Genannt wird immer wieder die fehlende Stabilität und ungelöste Frage der Langzeitarchivierung und -verfügbarkeit. Doch dieses Gefühl des Mangels, so die These des Beitrags, resultiert nicht aus noch nicht geklärten methodischen oder technischen Fragen, sondern beruht auf einer Fehleinschätzung der Natur digitaler Editionen, die in Analogie zum Druckmedium meist nur von ihrer Oberfläche her beurteilt werden. Mit einem Perspektivwechsel, der die Eigentümlichkeiten digitaler Editionen und die zugrundeliegende strukturellen und algorithmischen Komponenten ernst nimmt, ist indes vergleichbare Stabilität möglich, zumindest wenn man sich über das Dokumentenmodell und über die Form seiner technischen Realisierung verständigt.

Markup und Overlap - auf dem Weg zu einem konsolidierten Dokumentenmodell

Die Entwicklung und Nutzung der Markupsprache XML war von Anbeginn an begleitet von Kritik über die Unzulänglichkeit des hierarchischen OHCO Modell (DeRose et al. 1990) für die Repräsentation von Text. Trotz verschiedener Vorschläge konnte bis heute keine abschließende, alle spezifischen Kodierungsprobleme klärende Lösung gefunden werden. Nun hat das der Popularität von XML im Allgemeinen und der in diesem Feld maßgeblichen TEI im Besonderen nicht geschadet. Nach wie vor erfreut sich XML/TEI großer

Beliebtheit, auch wenn in jüngerer Zeit der Unterschied von TEI und XML betont wird (Cummings 2017). Das ist umso erstaunlicher, als es an alternativen Ansätzen nicht gemangelt hat (DeRoses 2004; Speerberg-McQueen 2007). Von MECS, GODDAG, TexMECS über LMNL bis zuletzt Text as a Graph (TAGML) entstanden Markup-Konzepte, die für sich in Anspruch nehmen und nahmen, XML und seine Beschränkungen zu überwinden. Gerade mit dem neuesten Konzept des *Text as a Graph* (Kuczera 2016; Dekker/Birnbaum 2017) scheint nach dem Selbstverständnis der Autoren nun endlich der Weg aus der Dauerkrise gewiesen. Doch auch dieses Modell, für das ein erster Serialisierungsentwurf vorliegt (Dekker et al. 2018), wirft erneut Fragen auf. Wenn es auch bestechend scheint und mehr Kodierungsflexibilität verheißt, ist doch fraglich, ob die Graphentheorie tatsächlich das Mittel der Wahl ist, um z.B. einem *autopoietischen* Textbegriff, wie (McGann 2016) ihn postuliert, Herr zu werden. Auch wenn die Graphentheorie als Modell von großem Nutzen sein kann, da sie uns hilft, mehr „clarity of thought about textuality“ (McGann 2016: 90) zu gewinnen, scheint es klug, in der Frage der „Textdefinition“ Vorsicht walten zu lassen und nicht aus dem Auge zu verlieren, dass es bei der Übersetzung des Textes in maschinenlesbare Form nicht nur darum geht, den „intellektuellen“ Textbegriff zu modellieren - als „model of“ (McCarty 2005) -, sondern auch und vor allem darum, das Textverständnis im Sinne eines „model for“ zu erweitern. So vermag man einen „Text“ zu konzipieren, von dem McCarty sagt, dass er einen „end maker“, keinen „end user“ benötigt und der es erlaubt, so etwas wie eine „digitale Pragmatik“ oder „digitale Hermeneutik“, vgl. u.a. (Scheuermann 2016), in die digitale Editorik einzuführen - in Anknüpfung an Konzepte von (Robinson 2003), (Shillingsburg 2006) oder (Gabler 2010).

Was aber von Anfang an bei der Diskussion um Overlap und die Unzulänglichkeiten von XML vernachlässigt wurde, ist, dass die Limitationen des Textmodells nicht unbedingt mit Limitationen der Serialisierung und der digitalen Technik in eins gesetzt werden können. So ist es zwar richtig, dass SGML und in der Folge XML mit dem OHCO Modell im Kopf entwickelt wurden, doch haben bereits (Renear et al. 1993) in ihrer Revision darauf abgehoben, dass im pragmatischen Sinne deskriptives Markup auch unabhängig von der OHCO These verwendet werden kann. Eben weil XML vor allem eine Syntax ist, war es letztlich immer wieder möglich, für „konforme“ Lösungen zu sorgen, wie die Vorschläge der TEI zu nicht-hierarchischen Strukturen (TEI Guidelines: Chap. 20) verdeutlichen, aber auch die Beiträge von Renear zum Konzept des „trojanischen Markup“ (Renear 2004) und zur XMLisierung von LMNL in CLIX (Renear 2004) oder xLMNL (Piez 2012). Der Grund lag auch darin, dass XML nicht nur in gut etablierte Strukturen der X-Familie eingebettet ist (XSD, XSLT, XQuery etc.), sondern auch allgemeiner die zentrale Document Object Modell (DOM)-Schnittstelle mit allen wichtigen Webelementen wie HTML bzw. XHTML (HTML 5), CSS oder Javascript teilt.

Multimodalität des Dokumentenmodells - nicht nur Markup

Trotz aller Experimente ist heute in der Praxis kaum strittig, dass die TEI und das in ihr entwickelte Dokumentenmodell und mit Abstrichen auch ihre Serialisierung in XML das Mittel der Wahl für digitale Editionen ist. Allerdings bleibt das Modell unvollständig, wenn man nicht auch die anderen Komponenten der Edition in die Überlegungen einbezieht. Wie das analoge Buch über die Rolle zum Kodex und über die Handschrift zum Druck gefunden hat, so muss auch das digitale Buch zu einer stabilen Struktur und Form finden, um nicht nur in der Wissenschaft, sondern auch in Gedächtniseinrichtungen wie Bibliotheken langfristig gesichert, reproduziert und als wissenschaftlich referenzierbares Objekt über Schnittstellen zur Verfügung gestellt und genutzt werden zu können. Dabei sind die Besonderheiten digitaler Dokumente bzw. ihre spezifische Dynamik bzw. Potentialität zu beachten (PDF z.B. erfüllt diese Kriterien nicht). So ist es für das Verständnis der digitalen Edition wichtig, in die Debatte um die Zu- oder Unzulänglichkeit bestimmter Markupsprachen auch das eine Edition verwirklichende Ensemble von Dateien und Funktionen einzubeziehen, deren logisches Zusammenspiel zu bestimmen und nicht nur nach dem, wie die TEI es nennt, „abstract model“ und dessen Serialisierungen zu unterscheiden, sondern auch Regelstrukturen, Präsentationsmodelle und differenzierte Metadatenformate als zum Verständnis notwendige Aspekte zu berücksichtigen. Editionen treten uns typischerweise als eine Kombination von Text mit Markup, Schemadatei, Stylesheets, Transformations- und sonstigen Skripten entgegen. Konkret handelt es sich um eine Reihe von Dateien (oder Datenströmen), wie z.B. .xml, .xslt, .xsd, .css oder .html, die zusammen ein funktionales Ganzes bilden, das als solches nicht nur die Stelle des physischen Dokumentes einnimmt, sondern auch die Grundlage der Langzeitarchivierung bildet. Dieser ganzheitlich betrachteten digitalen Edition ist eigentümlich, dass sie erst durch eine konkrete, meist nutzergesteuerte algorithmische Verarbeitungsanweisung nach dem klassischen EVA-Prinzip im Viewport oder Empfängersystem „verwirklicht“ wird, während ihre Persistenz in den in die Edition hineinkodierten und in ihren Darstellungsfunktionen niedergelegten Möglichkeiten, keineswegs aber in der sichtbaren Oberfläche liegt. Letztere reduziert sich zu einem Ausschnitt, der nur bedingt das gesamte Potential der Edition aufzeigen kann. Wenn diese kombinatorisch vollständig beschreibbaren Möglichkeiten der Präsentation, die zutreffend mit dem Begriff der Schnittstelle verbunden werden (Boot/Zundert 2011; Zundert 2018), den Kernbegriff der digitalen Edition konstituieren, resultieren daraus eine Reihe von praktischen und theoretischen Konsequenzen. Ein erster

wichtiger Schritt liegt in der Erkenntnis der Superiorität der Kodierungsgrundlage über die erzeugte angezeigte Oberfläche (Turska/Cummings/Rahtz 2016): „Data is the important Long-term Outcome“. Das heißt aber nicht, dass die Oberfläche gleichgültig wäre. Sie darf nur nicht, weil sichtbar, als das einzig wichtige, ja nicht einmal als für die Edition maßgebliche Layer begriffen werden. Die Oberfläche, Visualisierung, die Ausgabe, die Schnittstelle, allgemein das algorithmische Erzeugnis, können über die primäre, immer aber reduzierte und mit Blick auf die kombinatorischen Möglichkeiten ausschnittshafte Darstellungsfunktion hinaus ihrerseits eigenständige Produkte bzw. „Interpretationen“ sein, vgl. (Zundert 2018). Die Edition selbst sind sie aber nicht, denn eine digitale Edition kann man, streng genommen, nicht sehen. Entsprechend sind zum einen eine Reihe von Text- bzw. Dokumentelemente wie das Layout als eigenständiges bedeutungstragendes oder zumindest -beeinflussendes Phänomen als digitale „Textästhetik“ und als ein Ergebnis einer Funktion mit vielfältigen Parametern neu zu interpretieren (Stäcker 2019), zum anderen verändern sich Nutzungsszenarien etwa bei der Archivierung und Zitierbarkeit von Editionen, denn wenn die Oberfläche eine von mehreren Möglichkeiten ist, kann sie nicht ohne weitere Vorkehrung Gegenstand des Zitierziels sein. Auf anderer Ebene bedeutet es, dass der/die Autor/Autorin oder, vermutlich genauer, das Autorenteam ein genaues Verständnis auch der technischen Dimension des digitalen Textes haben muss, um seinen nicht nur natürlichen, sondern auch maschinellen „Leser“ zu erreichen, oder aber, dass die Autorintention die Schaffung von Möglichkeiten der digitalen Hermeneutik und Analyse einschließen muss. Eine wesentliche Dimension der digitalen Edition ist ferner ihre Verankerung im „Netz“. Daraus ergeben sich generell Anforderungen an ihre „Hypertextualität“, ihre „Schnittstellen“ (Zundert 2018; Stäcker 2019) und Fähigkeit, sich in das „semantic web“ zu integrieren (Ciotti/Tomasi 2016). Dazu zählt auch die unmittelbare Integration der genutzten „Forschungsdaten“, etwa der digitalen Faksimiles, die im Rahmen der *recensio* gesammelt und gesichtet wurden, so dass auf der Oberfläche ein transkludentes Ensemble entsteht, das auf *hypertextuellen* Strukturen aufbaut.

Es besteht die Hoffnung, dass mit dem Blickwechsel von dem zweidimensionalen sichtbaren Ergebnis auf die unsichtbare Potentialität der Edition sich das eher Proteushafte der Oberfläche der digitalen Edition auflöst und auch für die schon lange gärende Frage nach deren Persistenz und Nachhaltigkeit ein zufriedenstellender Ansatz gerade in ihrer, mit dem Motto der Tagung gesprochen: Multimodalität, gefunden werden kann. Der Beitrag möchte diesen Gedanken anhand von Beispielen weiter ausführen, um einen tragfähigen Begriff von einer persistenten digitalen Edition als einem funktionalen und organischen Ensemble von exakt definierbaren Komponenten zu entwickeln.

Bibliographie

Andrews, Tara L. / Zundert, Joris J. van (2018): *What Are You Trying to Say? The Interface as an Integral Element of Argument*, in: **Roman Bleier / Martina Bürgermeister / Helmut W. Klug / Frederike Neuber / Gerlinde Schneider (eds. / hrsg.): Digital Scholarly Editions as Interfaces.**

Norderstedt 3-33. urn:nbn:de:hbz:38-91064

Apollon, Daniel / Claire B  lisle / Philippe R  gnier (eds.) (2014): Digital critical editions. Topics in the digital humanities. Urbana: University of Illinois Press.

Boot, Peter/ Zundert, Joris van (2011): The Digital Edition 2.0 and The Digital Library: Services, not Ressources, in: Digitale Edition und Forschungsbibliothek. (Beitr  ge der Fachtagung im Philosophicum der Universit  t Mainz am 13. und 14. Januar 2011). Harrassowitz 2011: 141–52 (Bibliothek und Wissenschaft, 44) <http://hdl.handle.net/20.500.11755/c9e80904-8def-438e-a82b-80d4107b36ed>

Ciotti, Fabio / Tomasi, Francesca (2016): Formal Ontologies, Linked Data, and TEI Semantics, in: Journal of the Text Encoding Initiative. 10.4000/jtei.1480

Cummings, James (2017): Slides zum Vortrag gehalten auf der DH2017: <https://slides.com/jamescummings/teimyths> . Erscheint in K  rze in DSH.

Dekker, Ronald / David J. Birnbaum (2017): It's more than just overlap: Text As Graph. Presented at Balisage: The Markup Conference 2017, Washington, DC, August 1 - 4, 2017, in Proceedings of Balisage: The Markup Conference 2017. Balisage Series on Markup Technologies, vol. 19. <https://doi.org/10.4242/BalisageVol19.Dekker01>

Dekker, Ronald / Elli Bleeker / Bram Buitendijk, Astrid Kulsdom / David J. Birnbaum (2018): TAGML: A markup language of many dimensions. Presented at Balisage: The Markup Conference 2018, Washington, DC, July 31 - August 3, 2018, in Proceedings of Balisage: The Markup Conference 2018. Balisage Series on Markup Technologies, vol. 21. <https://doi.org/10.4242/BalisageVol21.HaentjensDekker01>

DeRose, Steven J. / Durand, David G. / Mylonas, Elli / Renear, Allen H.: What Is Text, Really?, in: Journal of Computing in Higher Education 1, Nr. 2 (Dezember 1990) 3–26. <https://doi.org/10.1007/BF02941632> .

DeRose, Steven J. (2004): Markup overlap: a review and a horse, in: Proceedings of Extreme Markup Languages. <http://xml.coverpages.org/DeRoseEML2004.pdf>

Deutsche Forschungsgemeinschaft (2015): F  rderkriterien f  r wissenschaftliche Editionen in der Literaturwissenschaft. Merkblatt der DFG. 2015. http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/foerderkriterien_editionen_literaturwissenschaft.pdf

Driscoll, Matthew James/ Elena Pierazzo (eds.) (2016): Digital scholarly editing: Theories and practices.

Cambridge, UK. <https://www.openbookpublishers.com/reader/483> .

Gabler, Hans Walter (2010): Theorizing the Digital Scholarly Edition, in: Literature Compass 7, Nr. 2 (Februar 2010) 43–56. <https://doi.org/10.1111/j.1741-4113.2009.00675.x> .

W3C (2017): HTML 5.2 Recommendation (14 December 2017): §9: XML Syntax. <https://www.w3.org/TR/html52/xhtml.html#xhtml>

Kuczera, Andreas (2016): Digital Editions beyond XML – Graph-based Digital Editions, in: Proceedings of the 3rd HistoInformatics Workshop on Computational History (HistoInformatics 2016). ceur-ws.org/Vol-1632/paper_5.pdf

McCarty, Willard (2005): Humanities computing. Basingstoke, Hampshire [u.a.].

McGann, Jerome J. (2014): A new republic of letters: memory and scholarship in the age of digital reproduction. Cambridge, Mass. [u.a.]: Harvard Univ. Press.

Piez, Wendell (2012): Luminescent: parsing LMNL by XSLT upconversion. Presented at Balisage: The Markup Conference 2012, Montr  al, Canada, August 7 - 10, in Proceedings of Balisage: The Markup Conference 2012. Balisage Series on Markup Technologies, vol. 8. <https://doi.org/10.4242/BalisageVol8.Piez01>

Renear, Allen / Mylonas, Elli / Durand, David (1993): Refining our Notion of What Text Really Is: The Problem of Overlapping Hierarchies. Final version, January 6. <http://cds.library.brown.edu/resources/stg/monographs/ohco.html>

Robinson, Peter (2003): WHERE WE ARE WITH ELECTRONIC SCHOLARLY EDITIONS, AND WHERE WE WANT TO BE, in: Jahrbuch f  r Computerphilologie. <http://computerphilologie.uni-muenchen.de/jg03/robinson.html>

Sahle, Patrick (2013): Digitale Editionsformen#: zum Umgang mit der   berlieferung unter den Bedingungen des Medienwandels. Schriften des Instituts f  r Dokumentologie und Editorik. Bd. 1-3. Norderstedt.

Scheuermann, Leif (2016): Die Abgrenzung der digitalen Geisteswissenschaften, in: Digital Classics Online vol. 2. <https://journals.ub.uni-heidelberg.de/index.php/dco/article/viewFile/22746/21865>

Shillingsburg, Peter L. (2006): From Gutenberg to Google: electronic representations of literary texts, Cambridge.

Speerberg-McQueen, C. M. (2007): Representation of overlapping structures: Proceedings of the 2007 Extreme Markup Languages conference. <http://conferences.idealliance.org/extreme/html/2007/SpeerbergMcQueen01/EML2007SpeerbergMcQueen01.html#id96245>

St  cker, Thomas (2018): ‚Von Alexandria lernen‘. Die Forschungsbibliothek als Ort digitaler Philologie. In Frauen – B  cher – H  fe: Wissen und Sammeln vor 1800 Women – Books – Courts: Knowledge and Collecting before 1800. Essays in honor of Jill Bepler. Hrsg. von Volker Bauer, Elizabeth Harding, Gerhild Scholz Williams

und Mara R. Wade. Wiesbaden: Harrassowitz 93–103.
<http://nbn-resolving.de/urn:nbn:de:tuda-tuprints-75938>

Stäcker, Thomas (2019): *Literaturwissenschaft und Bibliothek – eine Beziehung im digitalen Wandel*, in: Digitale Literaturwissenschaft. Metzler/Springer (im Druck, voraussichtl. 2019).

TEI Guidelines: *Chapter 20: Non-hierarchical Structures*. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/NH.html>

Turska, Magdalena / Cummings, James / Rahtz, Sebastian (2016): *Challenging the Myth of Presentation in Digital Editions*, in: Journal of the Text Encoding Initiative 9 (2016-2017). <http://journals.openedition.org/jtei/1453>