

Artikel aus:  
Zeitschrift für digitale Geisteswissenschaften

Titel:  
>A digital edition is not visible< – some thoughts on the nature and persistence of digital editions

Autor/in:  
Thomas Stäcker

Kontakt:  
[thomas.staecker@tu-darmstadt.de](mailto:thomas.staecker@tu-darmstadt.de)

Institution:  
Universitäts- und Landesbibliothek Darmstadt

GND:  
141905573


ORCID:  
0000-0002-1509-6960

---

DOI des Artikels:  
[10.17175/2020\\_005](https://doi.org/10.17175/2020_005)

Nachweis im OPAC der Herzog August Bibliothek:  
[882162004](#)

Erstveröffentlichung:  
16.10.2020

Lizenz:  
Sofern nicht anders angegeben 

Medienlizenzen:  
Medienrechte liegen bei den Autoren

Letzte Überprüfung aller Verweise:  
01.09.2020

GND-Verschlagwortung:  
[Edition](#) | [Elektronisches Publizieren](#) | [Informationsvermittlung](#) | [XML](#) |

Zitierweise:  
Thomas Stäcker: >A digital edition is not visible< – some thoughts on the nature and persistence of digital editions. In: Zeitschrift für digitale Geisteswissenschaften. Wolfenbüttel 2020. PDF Format ohne Paginierung. Als text/html abrufbar unter DOI: [10.17175/2020\\_005](https://doi.org/10.17175/2020_005).

Thomas Stäcker

›A digital edition is not visible‹ – some thoughts on the nature and persistence of digital editions

---

## Abstracts

Nach einer Phase des Experimentierens und Ausprobierens haben sich digitale Editionen weitgehend durchgesetzt und gelten als ernstzunehmende, ja bessere Alternative zu gedruckten Editionen. Mit TEI/XML steht zudem ein konsolidierter Standard für die Auszeichnung der relevanten strukturellen und semantischen Elemente einer Edition zur Verfügung. Doch nach wie vor begleitet diesen Prozess der Konsolidierung teils scharfe Kritik, vor allem mit dem Vorwurf, dass das auf XML basierende Markup ein Textmodell der *Ordered Hierarchy of Content Objects* (OHCO) begünstige, das zahlreichen Editionsproblemen nicht gerecht werde bzw. die Flexibilität des Editors einschränke. In der Folge sind vielfältige Versuche unternommen worden, diese Einschränkungen von XML zu überwinden, ohne dass sich jedoch eine der vorgeschlagenen Alternativen durchgesetzt hätte. Zugleich ist mit der Engführung der Kritik auf das scheinbar durch den XML Standard bedingte Textmodell aus dem Blick geraten, dass eine digitale Edition aus mehr besteht als lediglich einer XML Datei. Der Beitrag versucht aufzuzeigen, dass die einseitige Kritik an XML sich weitgehend auflöst, wenn die digitale Edition nicht nur als XML Datei, sondern als Ensemble ihrer Komponenten verstanden wird. Zugleich lässt sich daran aufzeigen, dass die digitale Edition, anders als ihre Kritiker behaupten, keinesfalls flüchtiger oder instabiler ist als ihr gedruckter Vorgänger. Das scheinbar Flüchtige löst sich auf, wenn sie nicht mehr aus der sichtbaren Oberfläche, sondern, ihrer algorithmischen Natur folgend, aus dem funktionalen Zusammenspiel aller Komponenten von Text, Struktur, Layout, Schnittstelle und Metadaten bestimmt wird.

After a period of experimentation and prototyping, digital editions are considered a common standard and a serious, quite often even a better alternative to printed editions. In addition the TEI/XML provides a well introduced standard for mark-up of all relevant structural and semantic elements of an edition. In spite of this process of consolidation the digital edition is still accompanied by harsh critique, particularly by objecting that mark-up leaning on XML fosters a text model of an *Ordered Hierarchy of Content Objects* (OHCO) that does not fit all editorial problems and limits the flexibility of the editor. As a consequence many attempts have been undertaken to overcome these limits of XML, but up to now without much success. By narrowing down the perspective, however, to problems of the text model seemingly caused by XML it was often overlooked that a digital edition consists of more than a XML file. This contribution attempts to show that the critique can be dissolved when the edition is viewed not merely as a XML file, but as an ensemble of its components. In doing so it can also be shown that other than its critiques maintain a digital edition is not less stable or persistent than its printed predecessor. The seeming fluidity of digital edition disappears if it is no longer determined by its visible surface, but according to its algorithmic nature by the interplay of its components of text, structure, layout, interface and metadata.

## 1. Introduction

After a period of experimentation and prototyping, digital editions have reached a state of consolidation in most Humanities' disciplines that entitles them to be considered no longer as special cases, but as normal editorial practice. A digital edition outstrips its predecessor by solving a couple of intricate editorial problems much better and by making results widely available via the internet. More recent publications such as Sahle, Apollon et al., Driscoll / Pierazzo and the guidelines of funding agencies such as DFG attest to this ongoing cultural shift.<sup>1</sup>

However, despite the progress digital editions have made in recent years, they still encounter serious criticism, particularly regarding the lack of stability and the still unsolved issue of long term archiving and accessibility, quite often leading to so-called *hybrid editions* that seemingly better preserve the editorial outcome against future decay and ensure its accessibility and quotability.

Yet this feeling of deficiency, so to speak, and concern about the longevity of many years labor seems to be rooted less in technical or methodological arguments, but in a profound misunderstanding of the nature of digital editions. Even if recent papers from the DH community suggest otherwise<sup>2</sup>, digital editions are still judged primarily by their appearance or their surface as it is visible to the reader and user, rather than on features that are more appropriate for explaining their particular character. Therefore, a change of perspective seems necessary to understand properly structures, functions and properties that are constitutive of digital editions in order to tackle the problems and concerns that accompany the transition from the analog to the digital edition.

---

<sup>1</sup> Sahle 2013; Apollon et al. 2014; Driscoll / Pierazzo 2016; guidelines of funding agencies such as DFG 2015.

<sup>2</sup> Turska et al. 2016, *passim*.

I argue in this paper that once the elements and functions of digital editions are sufficiently determined and properly understood, most of the problems that we encounter today with digital editions will disappear or at least be resolved much more easily. The shift in perspective I propose here illustrates shortcomings of the theoretical approaches of the previously dominant document and text model and offers a set of components or building blocks that help better understand the nature of digital editions and construct them in ways that ensure more reliability and instill trust in the digital format.

## 2. Markup and Overlap – towards a consolidated text and document model

Even though it was occasionally debated, digital editing started with markup.<sup>3</sup> The distinction between text in the narrower sense, on the one hand, and information about text called markup or in some cases annotation, on the other, proved to be the key invention for digital publishing in general and digital editing in particular. The invention, to be more precise, consists less in the inclusion of formatting or other signs in a text to produce the output of a rendition process (as is the case with early word processor software), but in the fact that markup is taken to be a distinct set of code that can be generally described by formal standards and that exists independently of the text at hand so that it can be shared and adapted to various requirements. The decisive invention, therefore, is the *general markup language*<sup>4</sup> which could be fixed as a set of rules in *Document Type Definitions* (DTD), later in other schema languages. Without this straightforward method that was easy to handle even by technical layman, digital editing would never have had the success it has had in recent years. As usual in such cases the breakthrough is less conditioned by the mathematical or intellectual beauty or soundness of the argument, but instead by various social factors and economic contexts that can be conceived of as a paradigm.<sup>5</sup> Critics like Schloen<sup>6</sup> fail to take these particular historical circumstances into account when he emphasises that the relational model devised by the mathematician Codd is superior to the markup model by the lawyer Goldfarb. Most influential was the development of SGML and its two offspring, HTML and XML.<sup>7</sup> Today XML is the de facto standard for digital editions and increasingly for digital publications, too, though in the case of the latter PDF encoded documents still prevail, thus defying the open markup concept by adhering to the analog paradigm in mimicking printed documents on the screen.

From the outset development and usage of the markup language XML and its most prominent application profile in the humanities the *Text Encoding Initiative* (TEI) was accompanied by criticism about the inappropriateness of the so-called *OHCO model*<sup>8</sup> for encoding complex textual phenomena and representing text properly. The main objection concerns the problem of overlap as XML has to be nested and does not allow for overlap. Cases of conflicting physical and logical structures, for instance, page versus division, are well known, but other structures may be incompatible as well, such as linguistic versus semantic encoding, e. g. connected semantic information that spreads over two sentences or paragraphs.

Notwithstanding various suggestions that have been proposed to solve these issues, no final solution could be provided that addresses all the issues raised by critics, and overlap issues continue to plague encoders,<sup>9</sup> not only those that adhere to the OHCO model. However, surprisingly enough, this has not diminished the popularity of XML in general and TEI in particular, and TEI/XML continues to be the medium of choice when it comes to digital editing, even if members of the TEI community more recently feel the need to emphasize that there is a difference between the TEI abstract model and its XML serialisation<sup>10</sup> In view of its acknowledged shortcomings, the perseverance of the TEI/XML model is most striking, especially since there is no lack of alternative solutions<sup>11</sup> to overcome the deficiencies of XML and the OHCO text model. Among them MECS, GODDAG, TexMECS, LMNL, CHOIR/OCHRE<sup>12</sup> all promise to transcend »the document paradigm«, or *Text as a Graph* (TAGML) that relies on a highly flexible graph model. What is it exactly that prevents all these solutions from replacing XML right away? For instance, the most recent concept *Text as a Graph*<sup>13</sup> seems to be an elegant and sound concept, promising a much greater deal of flexibility. A first serialization draft was put forward by Dekker et al.<sup>14</sup> and according to the authors this technique solves the problems that come with XML. However, along with older »convincing« solutions TAG shares the fate that while being »a good idea«, it only reaches a few very narrow specialist circles, while the great majority seems to be still satisfied with established XML solutions<sup>15</sup>, above all, because there are a lot of many useful tools available that work efficiently with XML. Above and beyond this, even if one might be

---

<sup>3</sup> Schloen 2014, *passim*.

<sup>4</sup> Goldfarb 1981; Sahle 2013, vol. 3, p. 142.

<sup>5</sup> Kuhn 1970, *passim*.

<sup>6</sup> Schloen 2014.

<sup>7</sup> DuCharme 1999, *passim*.

<sup>8</sup> DeRose et al. 1990, *passim*.

<sup>9</sup> Tennison 2008.

<sup>10</sup> Cummings 2019.

<sup>11</sup> DeRose 2004, *passim*; Sperberg-McQueen 2007.

<sup>12</sup> Schloen 2014, *passim*.

<sup>13</sup> Kuczera 2016; Dekker / Birnbaum 2017.

<sup>14</sup> Dekker et al. 2018.

<sup>15</sup> Cf. *Catalog of Digital Scholarly Editions*, Sahle et al. 2020.

convinced that TAG is better than XML, we do not know yet whether or not it really is suitable for solving all the issues that arise in the context of text encoding. The question remains whether the benefits of TAG cannot also be solved with XML, or whether changing is really such an advantage that it justifies replacing a well established infrastructure. In the long run, graph theory is equally insufficient to e. g. represent something like the *autopoietic textual concept* that McGann proposes.<sup>16</sup> Accordingly, even if graph theory as a way of modelling text may help gain more »clarity of thought about textuality«<sup>17</sup>, it seems wise to be cautious with respect to definite text models, and we should always be aware that converting text into a machine readable form does not, in and of itself, mean »understanding« the text. Markup is always a very limited way of making main structural features of a text explicit or of singling out particular entities. Consequently, suggesting a »model of«<sup>18</sup> text will always be restricted to very basic features while a »model for« approach seems to be more promising, as we can concentrate on functional characteristics that do not claim to explain the very nature of text, but make it operative for queries and manipulations. Text in this sense allows introducing a digital pragmatic or even a digital hermeneutics<sup>19</sup> into the digital edition. This approach opens up possibilities of working with and through digital texts in ways that are profoundly different from traditional approaches and former concepts of authorship by relating to ideas expressed by Robinson, Schillingsburg or Gabler, all of whom emphasize the social dimension of the digital edition.<sup>20</sup>

According to this pragmatic and functional sense of markup, the discussion about overlap and deficiencies of the OHCO text mode should, in the first instance, not be mistaken as a problem of the serialisation.<sup>21</sup> In other words, even if we agree that XML lends itself particularly well for representing the OHCO model, it is not identical to it. We could even implement the OHCO model in TAG. To be sure, SGML and XML have been developed in view of OHCO, but early reevaluations have already pointed out that in a pragmatic way descriptive XML markup can be used independently from the OHCO thesis.<sup>22</sup> Since XML is just a syntax, it proved to be very flexible in adapting to particular text models and the community over and over again managed to develop solutions for non-hierarchical structures conformant to the formal XML standard. This can be seen best in the development of the TEI<sup>23</sup>, and also in contributions such as Trojan markup<sup>24</sup> or XML conformant solutions of LMNL, i. e. CLIX<sup>25</sup> or xLMNL<sup>26</sup>. Even amongst its critics, XML is therefore a viable solution, not least because of its wide dissemination. A reason, which is often overlooked or neglected by proponents wishing to replace XML by new formats. What makes a strong case for using XML is the fact that it shares the *Document Object Model* (DOM)-Interface with all relevant web elements such as HTML/XHTML, CSS or Javascript. Accordingly, abandoning XML would mean to be forced converting data structures according to DOM once they are to be published on the web.

### 3. Not just markup – completing the document model

Currently TEI, its abstract model and its serialisation in XML are without a doubt the most important standard for digital editions. However, the model is incomplete when we confine ourselves to markup and ignore the other components and factors required to implement an edition. Just like the analog book that developed gradually from scroll to codex and from manuscript to print and came to obtain its particular standardized appearance and format, the digital book or edition also requires a »standardized« structure and form enabling it to be »shelved« by libraries and used by scholars seeking to apply digital methods. »Books« of that kind need to be reliably referenced, their content needs to be displayed in various forms, be it for reading, be it for sorting, or be it for other purposes. Machine-readability is a key prerequisite for digital editions and plays a crucial role for all kinds of applications, e. g. to perform queries, to analyse document content, or to link it to the semantic web. These kind of applications need to discern document structures, such as preface, introduction, abstract, main text, footnotes, bibliography, index and the like. Documents and data should be easily disseminated via standardized interfaces, concepts and entities be extracted and different versions synchronized. Users may wish to contribute to digital editions as social products and to engage in collaborative editorial projects. Seen against this background, it is imperative to leverage the full potential of the digital by avoiding rather traditional approaches that simply mimic the analog, e. g. by using PDF. This is why most theoretical works on digital editions share the conviction that they have to be considered entirely different from analog ones.<sup>27</sup> In view of this fundamental difference it is crucial to understand what actually is the gist of the digital edition and what are its pertinent components that allow researchers to devise a model that is stable enough to create reliable and consistent digital editions.

---

<sup>16</sup> McGann 2014, *passim*.

<sup>17</sup> McGann 2014, p. 90.

<sup>18</sup> McCarty 2005.

<sup>19</sup> Cf. Scheuermann 2016, p. 61.

<sup>20</sup> Robinson 2013; Robinson 2017; Schillingsburg 2006; Gabler 2010.

<sup>21</sup> For a reconsideration of OHCO as a text model see Sahle 2013, vol. 3, pp. 178.

<sup>22</sup> Renear et al. 1993.

<sup>23</sup> TEI guidelines, [chapter 20](#).

<sup>24</sup> DeRose 2004.

<sup>25</sup> DeRose 2004.

<sup>26</sup> Piez 2012.

<sup>27</sup> For instance Sahle 2013; Sahle 2016 or most of the contributions in Apollon et al. 2014.

In asking the question »What is a digital edition?«<sup>28</sup> yet again, we must shift the focus from markup theory to a broader view. A digital edition is more than an agreement on elements for markup, even though semantic or descriptive markup is an essential component. Besides text and markup at least two more aspects must be taken into account, namely both the manner in which encoded text is rendered (creating the output including layout and navigation) and the rules describing the model or structure of the document. Accordingly, when discussing advantages or disadvantages of markup, we should always be aware of this ensemble of interacting elements which together impacts the appearance and functions of the digital edition. Typically digital editions consist of files or data streams (e. g. XML, XSLT, XSD, CSS, javascript, HTML) containing text, images, objects, markup, rules, presentational or representational scripts, layout information, and descriptive and other metadata. What has to be shown in greater detail below is that at an abstract level we can discern a *semantic* or *structural model* for the phenomena to be encoded, a *process model* for the generation of the representation in a viewport, a *layout model* for the appearance of the document, a *communicative model*, so to speak, for interfaces and metadata providing descriptions about the edition. As mentioned, it is paramount to combine these different models conceptually and understand their interplay. Each makes an essential contribution to the digital edition as a whole and none can be discarded without losing a necessary element. To demonstrate the necessity of all elements working as an ensemble, one may think of two short textual paragraphs encoded with structural markup. Seemingly there is a great difference between

```
<div>
  <p>
    This is my first statement:
  </p>
  <p>
    Hello World!
  </p>
</div>
```

and

```
<div>
  <p>
    Hello World!
  </p>
  <p>
    This is my first statement:
  </p>
</div>
```

The first encoding seems to represent the right order while the second is plainly wrong. However, if we apply the following XSLT statements, the second version would be likewise correct with regard to the resulting output.

```
<xsl:value-of select="/div/p[2]" />
<xsl:value-of select="/div/p[1]" />
```

While nobody would encode text in the wrong order, the example provides a good idea about the interdependency of the semantics within the processing model<sup>29</sup> and demonstrates that we should not mix up the document model with the XML grammar.<sup>30</sup> To express this idea in another more subtle way, one must acknowledge »that XML has a semantics if – and only if – you also have a way of taking that representation and using it to produce something else, which, in practice, means having some means for transforming the XML data [...]«<sup>31</sup>

Another example relates to the choice issue raised by Dekker in his description of the TAG model. He argues that the order of TEI elements such as <abbr> and <expan> within <choice>,

---

<sup>28</sup> Sahle 2016.

<sup>29</sup> Sahle 2013, p. 109 and others (ib. FN 326) state that text is independent of and antecedes markup. Even if this is true for most cases, we have seen that in principle for XML this is no necessary condition: text and markup are interdependent.

<sup>30</sup> Flanders / Jannidis 2019, pp. 10.

<sup>31</sup> Ramsay 2019, p. 201.

if altered, results in a different document, even if the meaning is the same.<sup>32</sup> Again, this disregards the nature of XML documents whose appearance can be determined either by an explicit process model or by a schema file.<sup>33</sup> It makes no sense to insist on the order given in an XML file without taking the XSLT or schema file into account. A processor is not dependent on the given text order when it computes the file, but when using, e. g. a RelaxNG expression such as

```
<element name="choice">
  <interleave>
    <element name="abbr">
      <text/>
    </element>
    <element name="expan">
      <text/>
    </element>
  </interleave>
</element>
```

it »understands« that both orders are valid and conform to the same text model. To claim otherness of both versions is irrelevant or disregards the very nature of these kind of documents. In other words, it would not lead to a precedence of the one word order over the other, but should cause the processor to regard the elements as equivalent options, no matter what order is present. To be sure, a XML file without a schema can only represent a hierarchical structure being well formed, but once a schema is in place the particular semantics of the document can only be determined in light of the schema restrictions. Against this background it is understandable that Dekker et al. seek to eliminate all semantics from the application level (processing model) because this would contradict their argument.<sup>34</sup>

If we accept this holistic approach to the technical implementation of the digital edition it becomes evident that the edition is not realised but by an algorithmic process<sup>35</sup> that usually encompasses a couple of files such as .xml, .xslt, .xsd, .css or .js, each of which makes up a necessary component of the edition. These files are used for both the basis for the presentation of the edition in some kind of viewport and for its long term preservation. In other words, we have an ensemble of files that can and should be fully described. By drawing on established and transparent formats we are also able to ensure their long-term archival preservation. As already noted, not the surface, but »data is the important long-term outcome«. <sup>36</sup> Accordingly, one has to keep in mind that this kind of »edition« is by nature algorithmic.<sup>37</sup> It enacts the *input-process-output* model of all software applications. The XML document(s) are data that are being processed and the result is realized in a viewport of some kind. The result, the surface rendered by this process, is always derivative. Without the underlying data and functions it cannot be understood and should not be mistaken for the edition itself. For example, if there are scripts for displaying a version with an original and one for a normalized view encoded in the document, not just one of both constitute the edition, but the edition *is* the possibility to display both textual variants. The surface appearing on the viewport is just one of many other possibilities that, as mentioned, can be entirely described as a definite number of combinations and determined or produced by a finite set of statements and data information.

## 4. The Components of an Edition

Patrick Sahle has quite convincingly distinguished six different meanings of text in his monumental work on digital editions.<sup>38</sup> However, even if »the first task of any editor is to know what is meant by the terms document, text and work«<sup>39</sup>, because for identifying the constitutive parts of an edition, it is better to abstract from that sort of textual interpretation and focus on the basic technical conditions.

Text in this sense, therefore, is primarily determined by a sequence of code points and an application of markup. Sahle, while quoting the definition of markup put forward by Sperberg-McQueen: »By markup I mean all the information in a document other than the »contents« of the document itself, viewed as a stream of characters«<sup>40</sup>, questions that ASCII or Unicode are sufficient to characterize this kind of text. He objects: »dass er genau mit einem eng begrenzten, durch technische Bedingungen

---

<sup>32</sup> Dekker / Birnbaum 2017, p. 74.

<sup>33</sup> Ramsay 2019, *passim*.

<sup>34</sup> Dekker / Birnbaum, p. 73.

<sup>35</sup> Sahle 2013, vol. 2, p. 161; vol. 3, p. 108.

<sup>36</sup> Turska et al. 2016, p. 4.

<sup>37</sup> Sahle 2013, vol. 3, p. 148.

<sup>38</sup> Sahle 2013, vol. 3, p. 8.

<sup>39</sup> Robinson 2017, p. 7.

<sup>40</sup> Sperberg-McQueen 1991, p. 35.

bestimmten Grund-Zeichenraum zusammenfallen sollte, ist wenig plausibel«<sup>41</sup>. To be sure, editorial theory is related to semiotics or bibliography, and it is important to gain a clear idea of what kind of editorial principle an edition pursues. However, even though there are many meanings of text, as demonstrated by Sahle and others, Unicode is the technical precondition of all of them, and despite Sahle's qualms it makes good sense to take the Unicode character set to be a basic component of a digital edition and a prerequisite for long term archival. The Unicode standard limits the way text can be read by readers and by machines and the Unicode codepoint set is at the core of what text technically is.<sup>42</sup> Even if easily overlooked, digital text conforms to the *input-process-output* principle, too, depending on a rendering engine and a graphic representation of the letter usually described in a font. Text as a sequence of Unicode codepoints forms the basis of all editorial texts, while the code and graph together form the rendered output on the viewport.<sup>43</sup>

In addition or extension to the text proper further media objects can be included into the edition such as images, objects (3D), audio or video. These are subsumed under »text« as they are treated on the content level. They accompany the edition as research data and have both illustrative and verifying functions as in the case of digital facsimiles providing a view of the sources or testimonies on the basis of which the edition was compiled. For most of these, standard formats for long term archiving are available, e. g. TIFF as a robust standard for still images. For some others the situation is still unclear, e. g. for 3D representations.

In addition to the Unicode content comes markup, which should be defined by a schema file. Supplying an edition with a customized schema file allows for explaining the usage of the markup applied to the text in a machine-readable form. The schema file operates as a set of rules for the edition or as a documentary file that allows computers to »understand« the encoding. The schema file, e. g. an XSD or RelaxNG, can also be used to enrich the edition with semantic meaning.<sup>44</sup>

The difference between text and markup is sometimes blurred in view of codes such as *carriage return* or *line feed* that can likewise be encoded by markup or by Unicode codepoints, but in practice markup is preferred to formatting codepoints. Markup and text are closely related and the possible usage of both parts are encoded in the schema file. These three components constitute the »textual«, or more precisely, the structural level of the edition. For the sake of long term archiving we have to claim that these three components are present and discernable:

1. a sequence of codepoints (Unicode),
2. structural or descriptive markup (no matter, if inline or stand-off), and
3. a schema defining elements, attributes, and structures permitted in the edition. Usually, there is also a TEI/XML file and a customization of the TEI (ODD, RelaxNG).

The next important component is made up by the processing requirements of the edition. As a whole it can be described by a processing model as suggested e. g. by the TEI.<sup>45</sup> Yet this processing model is rather abstract and only provides a general description independent of its actual implementation in some programming language. Typical for XML is XSLT or XQuery. Other programming or scripting languages are possible, but not recommendable as they run the risk to encode functionalities in code that is proprietary and leans on a particular software. The W3C standards XSLT and XQuery are presumably less incumbent on obsolescence and due to their particular XML logic easier to integrate in editorial projects.

One last component, the layout, is often overlooked. Some seem to assume that layout plays no, or only a minor, role for digital texts and that we have to be clear about the semantics in the first place. But this would disregard the affordances of an aesthetic digital presentation. The history of print has shown the relevance and influence of aesthetic presentations<sup>46</sup>, whose semantics are often not really determinable. It makes a difference whether a page is set in Roman or Gothic font, and it makes a difference whether there is sufficient space in between the lines and at the border, even when the semantics are not fully clear. A site well designed with pleasant fonts, colors, and a printing or rather viewing area that is easy to survey, to navigate, or agreeable to read makes the edition delightful for the user and offers a better usability and understanding of the meaning. What is missing today is a new printers' art, so to speak, for digital presentations, that is, an art that is in a way very close to media design, but focuses on the beauty and usability of digital publications. This component encompasses all of the techniques to show or use the edition at the viewport. As a matter of course, what we get depends on the output format. The basic output format for digital editions is HTML, but there may be also other formats such as PDF, EPUP or even LaTeX, DOCX, ODT or similar outputs. Not all of these formats are suitable for including all of the many possible outputs a digital edition is able to generate. For instance, a PDF is perhaps not a good choice for conveying complex navigational logic in the display of different text versions, but provides

---

<sup>41</sup> Sahle 2013, vol. 3, p. 238.

<sup>42</sup> For the development of ASCII and Unicode see Sahle 2013, vol. 3, p. 118.

<sup>43</sup> Stäcker 2018.

<sup>44</sup> See TEI description of <equiv>. <equiv> is used for »naming the underlying concept of which the parent is a representation«. Cf. TEI guidelines, chapter 22.

<sup>45</sup> TEI guidelines, chapter 22.

<sup>46</sup> McKenzie 1999, p. 4.

the best choice for just simply reading pages on a computer screen. The same holds true for EPUB which is a good choice for smartphones or tablets, but fails to serve applications that attempt to extract structured information. Sahle observed that the current change of technology often leads to a loss of aesthetic vigor<sup>47</sup>, and McKitterick pointed out that in view of the material evidence »it is not a straightforward or speedy translation from original to screen«. <sup>48</sup> Therefore, it seems advisable to pay more attention to design issues and appropriate presentations.

## 5. Integration of the Components

In view of the peculiar nature of the digital edition, Price raised the question what the best term might be to describe it properly.<sup>49</sup> Besides edition he suggested project, database, archive, thematic research collection and arsenal, and indicated a preference for the latter. Whatever term one may choose, it becomes clear from his discussion that the classical term *edition* is no longer applicable and, if employed, must be re-defined. The reason lies not only in its volatility and fluidity, discussed above, and which is explained by its algorithmic nature, but also in its lack of integrity and various degrees of coherence.<sup>50</sup> However, in practice the changing states of coherence never lead to a complete dissolution of the editorial text, and it holds true that »whole of all its possible variant readings and interpretations makes up a virtual unity identical with itself. Text in this view is not an arbitrary unity, for if it were seen as such, no text would differ from any other.«<sup>51</sup>

Sometimes it seems difficult to distinguish the editorial core from additional material, or to evaluate the quality and persistence of resources that are linked to the edition. For instance, do the digital facsimiles of sources belong to the edition proper? Is an external resource, such as an authority file record, to which the edition provides a link part of the edition? Hypertext is a powerful tool, but brings about conceptual challenges. Digital editions are always net or hyper editions. If the various parts of the edition need not be at one place, but can be assembled by hyperlinks, issues of permanent availability and responsibility arise. According to the principle of transclusion,<sup>52</sup> viewport and editorial components may be at various locations, but if the editor or editorial team has no control over the distributed resources, reliability and quality becomes jeopardized as resources may disappear or change their meaning over time. In general, there are two ways to address this problem. The best solution is copying all resources to the edition. This allows for the exact mirroring of the state of knowledge the editor possessed at a certain point in time. If this is not possible, external resources should be linked, documented, and displayed in a way that the external provenance is visible and that readers or users are aware of possible alterations for which the editor is not responsible. Of course, some resources or digital sources are more reliable than others, but this belongs to a new kind of digital literacy the user needs in order to evaluate the resource properly.

All resources bundled into an edition should be carefully documented. This documentation must be integrated into the edition itself and makes up a central component describing resources, files, metadata, and the structure as well as processing instructions for rendering the digital edition. This assembly as a vivid organism can be viewed as what Robinson ties to the notion of »work«. <sup>53</sup> How this is accomplished is not standardized. Some include this information in a TEI document, others use METS<sup>54</sup>, some use other means for this purpose. To ensure the integrity of the edition, all internal and external resources must be listed here so as to determine what is part of the edition and how the components may interact.

## 6. The edition as interface

If the particular nature of the digital edition is seriously considered as being entirely different to the analog or printed edition, some basic features of the printed edition that are no longer operative with the digital one must be reconsidered: Consequently, the author, publisher, or other contributor of an edition must be aware of the functional and dynamic opportunities a digital edition has to offer in order to serve both the natural and the mechanical »reader«. The authorial intention, so to speak, has to take into account not only new ways of digital hermeneutics and analysis, but also the fact that a digital edition is always an internet edition and, therefore, part of a larger network. Hence requirements occur with respect to its hypertextuality and interfaces<sup>55</sup> or their suitability to be integrated into the semantic web.<sup>56</sup> As already mentioned, research data such as facsimiles of primary sources collected in the course of compiling sources (*recensio*) should be included either by downloading or linking.

---

<sup>47</sup> »Verlust von ästhetischer Kraft«, Sahle 2013, vol. 3, p. 123 and FN 361.

<sup>48</sup> McKitterick 2013, p. 2.

<sup>49</sup> Price 2009.

<sup>50</sup> Henrich 2005, pp. 21-25

<sup>51</sup> Buzetti / McGann 2006, p. 66.

<sup>52</sup> Cf. Wikipedia: [Transclusion](#).

<sup>53</sup> Robinson 2013, pp. 28-30.

<sup>54</sup> For instance see [METS file](#) of Wolfenbütteler Digital Library 2020.

<sup>55</sup> Andrews / Zundert 2018; Stäcker 2020.



One of the most challenging issues is quotation. Because there is no stable surface or carrier material like sheets of paper that allow for reliable pagination or numbering of lines, a different technique must be applied. This is perhaps one of the most fundamental changes in adopting the digital edition. Previous citation mechanisms are no longer viable; the citation itself changes its very nature. To come to grips with this problem, two basic principles must be acknowledged: first that a digital edition is always also an internet edition addressed by URLs, and second that in order to reconstruct a particular view, instructions and data are necessary to reproduce this same view must be available. In other words, the process of quoting is nothing else than an instruction expressed by an URL that encompasses or activates all information necessary to reconstruct a particular view or repeat a particular behavior of the edition.

Consequently, by quoting a digital edition we no longer refer to some point in the text as was the case with the printed edition, but by means of an URL we activate a mechanism that generates the view or version to be quoted. Roughly speaking, this URL refers to an editorial REST interface.<sup>57</sup> It is employed to submit, e. g. a GET statement evoking a particular response (view). Contrary to using an URL as basically a locator that identifies documents or fragments, a REST interface is a suitable means of communication with digital editions as it corresponds best to their algorithmic nature. Accordingly, a quotation is to be regarded as a particular communication act and a REST interface or more generally an interface<sup>58</sup> that allows for precisely evoking the view one wishes to reference. Even if it is limited to a subset of possible views, a good example of such a RESTful interface in connection with an edition can be explored at the [Carl Maria von Weber edition](#) at Paderborn, which uses the Swagger/OpenAPI software.

We are now able to better comprehend how the digital edition is closely tied to the concept of the interface and how the interface makes up a core feature of the digital edition,<sup>59</sup> how the interface is itself a way of ›interpreting‹ the edition,<sup>60</sup> and how to communicate with it (e. g. reading, extracting, ordering, sorting, or querying). For instance, an ›interface URL‹ may create a view that is produced by applying particular XML and XSLT or XQuery files. It is paramount that this kind of response is not dependent on proprietary software frameworks but is an integral part of the edition itself. This allows for copying the edition as a whole, similar to a book that can be moved from one place to another without losing any of its functionalities. Until now there has been no standard for setting this kind of ›quotation‹ to work, but we can already discern three key ingredients, i. e. the values or data (XML), the functions (XSLT, XQuery), and the parser (e. g. saxon, Xalan). Since the parser is standardized, we can ignore it here. Accordingly, it should suffice to quote the XML and XSLT or XQuery file to create a view. A possible URL might look like the following:

```
http://my-digital-edition.de/views?value="document1"&function=display1
```

However, this kind of interface functionality has to be stable. As is well known, URLs tend to disappear or change. To address this problem it is necessary to assign persistent identifiers to the edition or parts thereof. This can be achieved by using a PURL, DOI, Handle, URN or other mechanism ensuring that the ›quoted‹ statement always retrieves the same view. It should be noted, however, that the main purpose of PIs is identifying by naming, not locating. Thus a persistent identifier can be assigned to several items of the same edition. Locating functionality is provided by a resolving mechanism such as <https://doi.org/> for DOIs or <https://nbn-resolving.org/> for URNs, but can also be achieved by any search engine.

What is more difficult to achieve is to persistently address views for whose creation parameter values are required, as in the case above. The typical PIs such as DOI or URN are designed for static objects, and it seems hardly feasible or sensible to attach a static PI to every possible parameter view the edition has to offer. What is required, therefore, is a PI solution for reliably and persistently reproducing interface instructions. This type of PI is missing so far. OpenURL comes close to what is required, but is employed for a different purpose. The Handle mechanism of ›templating‹ may work to some extent,<sup>61</sup> but has some drawbacks. Another approach could be to extend the fragment identifier schemes<sup>62</sup> by query instructions.

In turn, the documents should be equipped with IDs to ensure granular citation of parts to be quoted. For instance, our example:

---

<sup>56</sup> Ciotti / Tomasi 2016, *passim*.

<sup>57</sup> Cf. Wikipedia: [Representational State Transfer](#).

<sup>58</sup> For the time being REST seems to be the most suitable kind of interface for this purpose (see Lanthaler 2014), but other solutions like SOAP, WSDL or RPC are not excluded.

<sup>59</sup> Boot / Zundert 2011, *passim*; Andrews / Zundert 2018.

<sup>60</sup> Andrews / Zundert 2018, p. 21.

<sup>61</sup> Cf. Handle.net (Version 9), 2018.

<sup>62</sup> Cf. Fenner 2014.

```
<div xml:id="id001">
  <p xml:id="id002">
    This is my first statement:
  </p>
  <p xml:id="id003">
    Hello World!
  </p>
</div>
```

could be quoted by

```
http://my-digital-edition.de/views?value="document1#id002"&function=display1
```

or by using a query instruction:

```
http://my-digital-edition.de/views?value="//p" &function=display1
```

As a matter of course, quoting via an interface can be further differentiated as to the kind of usage the user makes of the edition. Interfaces are employed not only by way of reading the whole or parts of the edition, but also of searching or working on the edition. Searching an edition can be regarded as a particular kind of machine-reading. A search action ideally expressed in a formal query grammar (e. g. CQL, XPath, SPARQL etc.) or simply as a full text search results in a list of hits pointing to pertinent passages or offering snippets for inspection. The hits themselves as single items or as a set can also be viewed as quotable text parts. On the one hand, provision has to be made that the edition can be interrogated by such a query. Input, on the other hand, can be brought to the edition by any format, as long as it conforms to the data model specified in the schema expressing the model and as long as it is suitable for a lossless conversion to the targeted XML/TEI format. Therefore, even if it is not to be recommended it is nonetheless possible to use other editing software that employs formats other than XML/TEI, for instance *Word* when equipped with suitable templates. The sense and relevance of the interfaces component consists exactly in that it functions according to a formalized data model allowing for conversion of data into whatever form is needed. Conversely, serving an interface means to render all proprietary modelling without loss of its integrity as conformant to the edition model manifest in the interface.

## 7. Conclusion

As demonstrated above, the digital edition does not have much in common with the traditional printed one. Its dynamic or algorithmic nature must be taken seriously. This means that we have to abandon the two-dimensional form of the printed book as authoritative in order to open ourselves to the particular character of the digital one. We go wrong when we try to compare digital and printed editions only according to their surface features. There can be no stability and persistence of the surface of the digital edition, and even if some scholars claim to be able to persistently archive an output format such as PDF/A, they do not do justice to, or even in some cases mutilate, the digital edition when they reduce it to just one of many possible output formats, data and document representations, or search options.

A digital edition as such is not visible, but all its components can be fully described and views reliably reproduced. By acknowledging the algorithmic nature of the digital edition, it loses its protean character and becomes as stable and reliable as the printed edition used to be over the centuries. All of its components that we have identified as making up the edition — text and markup, editorial functions, layout for output or input, (descriptive, structural, administrative or technical) metadata and interfaces — are for the most part already standardized by agencies such as W3C, leading national libraries, or are in the process of being standardized in order to provide a reliable framework for the centuries to come. Perhaps it is quite the opposite. By its very nature of allowing easy and lossless dissemination the digital edition is even more stable than a paper based editions ever was, because in stark contrast to digital editions print editions are prone to material decay.

## Bibliographic references

- Tara Lee Andrews / Joris Job van Zundert: »What Are You Trying to Say?« The Interface as an Integral Element of Argument. In: Digital Scholarly Editions as Interfaces. Ed. by Roman Bleier / Martina Bürgermeister / Helmut Werner Klug / Frederike Neuber / Gerlinde Schneider. Norderstedt 2018, pp. 3–33. (= Schriften des Instituts für Dokumentologie und Editorik, 12) URN: [urn:nbn:de:hbz:38-91064](https://nbn-resolving.org/urn:nbn:de:hbz:38-91064) [Nachweis im GBV]
- Digital critical editions. Topics in the digital humanities. Ed. by Daniel Apollon / Claire Bélisle / Philippe Régnier. Urbana, IL et al. 2014. [Nachweis im GBV]
- Peter Boot / Joris Job van Zundert: The Digital Edition 2.0 and The Digital Library: Services, not Ressources. In: Digitale Edition und Forschungsbibliothek. Ed. by Christiane Fritze / Franz Fischer / Patrick Sahle / Malte Rehbein. (Fachtagung, Mainz 13.–14.01.2011) Wiesbaden 2011, pp. 141–152. (= Bibliothek und Wissenschaft, 44) Handle: [20.500.11755/c9e80904-8def-438e-a82b-80d4107b36ed](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-20500-11755-c9e80904-8def-438e-a82b-80d4107b36ed) [Nachweis im GBV]
- Dino Buzzetti / Jerome McGann: Critical Editing in a Digital Horizon. In: Electronic Textual Editing. Ed. by Lou Burnard / Katherine O'Brien O'Keefe / John Unsworth. New York, NY 2006, pp. 53–73. [Nachweis im GBV]
- Fabio Ciotti / Francesca Tomasi: Formal Ontologies, Linked Data, and TEI Semantics. In: Journal of the Text Encoding Initiative 9 (2016–2017). DOI: [10.4000/jtei.1480](https://doi.org/10.4000/jtei.1480)
- HANDLE.NET® (Version 9). Technical Manual. Ed. by Corporation for National Research Initiatives. Version 9, Preliminary edition from June 2018. PDF. Handle: [20.1000/113](https://nbn-resolving.org/urn:nbn:de:hbz:5:1-201000113)
- James Cummings: A world of difference: Myths and misconceptions about the TEI. In: DSH 34 (2019), i. 1, pp. 58–79. Article from 14.12.18. DOI: [10.1093/llc/fqy071](https://doi.org/10.1093/llc/fqy071)
- Ronald Dekker / David Jonathan Birnbaum: It's more than just overlap: Text As Graph. In: Proceedings of Balisage: The Markup Conference 2017. (Conference, Washington, DC, 01.–04.08.2017) Red Hook, NY 2017. (= Balisage Series on Markup Technologies, 19) DOI: [10.4242/BalisageVol19.Dekker01](https://doi.org/10.4242/BalisageVol19.Dekker01)
- Ronald Dekker / Elli Bleeker / Bram Buitendijk / Astrid Kulsdom / David Jonathan Birnbaum: TAGML: A markup language of many dimensions. In: Proceedings of Balisage: The Markup Conference 2018. (Conference, Washington, DC, 31.07.–03.08.2018) Red Hook, NY 2018. (= Balisage Series on Markup Technologies, 21) DOI: [10.4242/BalisageVol21.HaentjensDekker01](https://doi.org/10.4242/BalisageVol21.HaentjensDekker01)
- Steven Joseph DeRose / David G. Durand / Elli Mylonas / Allen H. Renear: What Is Text, Really? In: Journal of Computing in Higher Education 1 (1990), i. 2, pp. 3–26. [Nachweis im GBV]
- Steven Joseph DeRose: Markup overlap: a review and a horse. In: Proceedings of Extreme Markup Languages. (EML: 4, Montréal, 02.–06.08.2004) Montréal 2004. PDF. [online]
- Förderkriterien für wissenschaftliche Editionen in der Literaturwissenschaft. Ed. by Fachkollegium Literaturwissenschaft der Deutschen Forschungsgemeinschaft. Bonn 2015. PDF. [online]
- Digital scholarly editing: Theories and practices. Ed. by Matthew James Driscoll / Elena Pierazzo. Cambridge 2016. [online] [Nachweis im GBV]
- Bob DuCharme: XML: The annotated specification. Upper Saddle River, NJ 1999. [Nachweis im GBV]
- Manfred Fenner: Fragment Identifiers and DOIs. In: martinfenner.org. Blog post from 02.08.2014. [online]
- Julia Flanders / Fotis Jannidis: The shape of data in Digital Humanities. Modeling texts and text-based resources. London et al. 2019. [Nachweis im GBV]
- Hans Walter Gabler: Theorizing the Digital Scholarly Edition. In: Literature Compass 7 (2010), i. 2, pp. 43–56. [Nachweis im GBV]
- Charles F. Goldfarb: A generalized approach to document markup. In: Proceedings of the ACM SIGPLAN SIGOA Symposium on Text Manipulation. (Symposium, Portland, OR, 08.–10.06.1981) New York, NY 1981, pp. 68–73. [Nachweis im GBV]
- Andreas Henrich: Spurenlesen: Hyperlinks als kohärenzbildendes Element in Hypertext. München 2005. URN: [urn:nbn:de:bvb:19-30544](https://nbn-resolving.org/urn:nbn:de:bvb:19-30544)
- Andreas Kuczera: Digital Editions beyond XML – Graph-based Digital Editions. In: Proceedings of the 3rd HistoInformatics Workshop on Computational History. Ed. by Marten Düring / Adam Jatowt / Johannes Preiser-Kappeller / Antal van Den Bosch. (HistoInformatics: 3, Krakow, 07.11.2016) Aachen 2016. (= CEUR workshop proceedings, 1632) [online] [Nachweis im GBV]
- Thomas Samuel Kuhn: The structure of scientific revolutions. 2. edition, enlarged. Chicago 1970. [Nachweis im GBV]
- Markus Lanthaler: Third generation Web APIs. Bridging the gap between REST and Linked Data. Graz 2014. [online]
- Willard McCarty: Humanities computing. Basingstoke et al. 2005. [Nachweis im GBV]
- Jerome John McGann: A new republic of letters: memory and scholarship in the age of digital reproduction. Cambridge, MA et al. 2014. [Nachweis im GBV]
- Donald Francis McKenzie: Bibliography and the sociology of texts. Cambridge 1999. [Nachweis im GBV]
- David McKitterick: Old books, new technologies: the representation, conservation and transformation of books since 1700. Cambridge et al. 2013. [Nachweis im GBV]
- Wendell Piez: Luminescent: parsing LMNL by XSLT upconversion. In: Proceedings of Balisage: The Markup Conference 2012. (Conference, Montréal, 07–10.08.2012) Red Hook, NY 2012. (= Balisage Series on Markup Technologies, 8) DOI: [10.4242/BalisageVol8.Piez01](https://doi.org/10.4242/BalisageVol8.Piez01)
- Kenneth M. Price: Edition, project, database, archive, thematic research collection: what's in a name? In: Digital Humanities Quarterly 3 (2009), i. 3. [online]
- Stephen Ramsay: Where semantics lies. In: The shape of Data in Digital Humanities: modeling texts and text-based resource. Ed. by Julia Flanders / Fotis Jannidis. London et al. 2019, pp. 197–203. [Nachweis im GBV]
- Allen Renear / Elli Mylonas / David Durand: Refining our notion of what text really is: the problem of overlapping hierarchies. In: [cds.library.brown.edu/resources/stg/monographs/ohco.html](https://cds.library.brown.edu/resources/stg/monographs/ohco.html). Providence, RI 1993. [online]
- Peter Robinson: The concept of the work in the digital age. In: Ecdotica 10 (2013), pp. 13–41. [Nachweis im GBV]
- Peter Robinson: Some principles for making collaborative scholarly editions in digital form. In: Digital Humanities Quarterly 11 (2017), i. 2. [online]
- Patrick Sahle: Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Schriften des Instituts für Dokumentologie und Editorik. 3 volumes. Norderstedt 2013. [Nachweis im GBV]
- Patrick Sahle: What is a scholarly digital edition? In: Digital scholarly editing: theories and practices. Ed. by Matthew James Driscoll / Elena Pierazzo. Cambridge 2016, pp. 19–40. [Nachweis im GBV]
- A catalog of Digital Scholarly Editions. Ed. by Patrick Sahle / Georg Vogeler / Jana Klinger / Stephan Makowski / Nadine Sutor. In: [digitale-edition.de](https://digitale-edition.de). V.4.021 2020ff from 07.09.2020. [online]
- Leif Scheuermann: Die Abgrenzung der digitalen Geisteswissenschaften. In: Digital Classics Online 2 (2016), i. 1, pp. 58–67. DOI: [10.11588/dco.2016.1.22746](https://doi.org/10.11588/dco.2016.1.22746) [Nachweis im GBV]
- David Schloen / Sandra Schloen: Beyond Gutenberg: Transcending the document paradigm in Digital Humanities. In: Digital Humanities Quarterly 8 (2014), i. 4. [online]
- Peter L. Shillingsburg: From Gutenberg to Google: Electronic representations of literary texts. Cambridge 2006. [Nachweis im GBV]
- Christopher Michael Sperberg-McQueen: Text in the Electronic Age: Textual Study and Textual Study and Text Encoding, with Examples from Medieval Texts. In: Literary and Linguistic Computing 6 (1991), i. 1. DOI: [10.1093/llc/6.1.34](https://doi.org/10.1093/llc/6.1.34)

Christopher Michael Sperberg-McQueen: Representation of overlapping structures. In: Proceedings of the 2007 Extreme Markup Languages conference. (EML: 7, Montréal, 07.–10.08.2007) Montréal 2007. [\[online\]](#)

Thomas Stäcker: »Von Alexandria lernen«. Die Forschungsbibliothek als Ort digitaler Philologie. In: Frauen – Bücher – Höfe: Wissen und Sammeln vor 1800 Women – Books – Courts: Knowledge and collecting before 1800. Essays in honor of Jill Bepler. Ed. by Volker Bauer / Elizabeth Harding / Gerhild Scholz Williams / Mara R. Wade. Wiesbaden 2018, pp. 93–103. URN: <urn:nbn:de:tuda-tuprints-75938> [\[Nachweis im GBV\]](#)

Thomas Stäcker: Literaturwissenschaft und Bibliothek – eine Beziehung im digitalen Wandel. In: Digitale Literaturwissenschaft. (DFG-Symposium Digitale Literaturwissenschaft, Villa Vigoni, 09.–13.10.2017) Stuttgart 2020. (in publication)

TEI Guidelines. P5: Guidelines for Electronic Text Encoding and Interchange. Chapter 20: Non-hierarchical Structures. Ed. by Text Encoding Initiative. Version 4.0.0 from 13.02.2020. [\[online\]](#)

Jenni Tension: Overlap, containment and dominance. In: [jenitension.com](http://jenitension.com). Blog post from 06.12.2008. [\[online\]](#)

Magdalena Turska / James Cummings / Sebastian Rahtz: Challenging the myth of presentation in digital editions. In: Journal of the Text Encoding Initiative 9 (2016–2017). DOI: [10.4000/jtei.1453](https://doi.org/10.4000/jtei.1453)

Representational State Transfer. In: Wikipedia. Die freie Enzyklopädie. Lexikonartikel vom 22.07.2020. [\[online\]](#)

Transclusion. In: Wikipedia. Die freie Enzyklopädie. Lexikonartikel vom 22.08.2020. [\[online\]](#)

Wolfenbütteler Digital Library. In: [diglib.hab.de](http://diglib.hab.de). Ed. by Herzog August Bibliothek. Wolfenbüttel 2020. [\[online\]](#)