

# Proto-editions: Historians and the "Something between digital image and digital scholarly edition"

**Vogeler, Georg**

georg.vogeler@uni-graz.at  
Universität Graz, Austria

This position paper suggests identifying a specific form of scholarly edition coming into existence by the digital transformation of this scholarly practice, the "proto-edition". It starts with the obvious observation that digital reproduction has become everyday practice in the archives. It is also consensus that simple publication of visual surrogates is not sufficient as a scholarly edition. However, the community of scholarly editors is aware of the fact that digital methods extend the scope of editorial work towards the archive and digital reproduction. It has even sparked a discussion on terminology (Price 2009, Sahle 2007, van Hulle 2009, Eggert 2017, Dillen 2019): how much of the digital scholarly edition is a database, an archive, an exhibit, a project?

For philological work, the difference between the visual reproduction and the transcription is obvious. For historians the situation is different. Text has a different quality there: it is not the main subject of research, but a medium to convey information about human activities and beliefs in the past. They consider texts as information carrier for historical observations. This can, of course be information on specific wordings, but most of it is information conveyed through the text. The text itself is somehow transparent. This has led Kaplan & di Leonardo (2017) to the interpretation that archival records are just a form of data representation that can be converted into digital forms, a "simple" re-documentation (Pédaque 2007): census records, a land register, account books etc. should be converted into those databases which the clerks and officials writing them down in the past would have used, if technology would have been available.

This is a simplified view of historians and, in general, humanities work. DH is aware that tracing scholarly interpretations to their source is crucial, and the modeling work influences our perception of the "original" information significantly. However, it is a correct view that historians create databases from their sources. Manfred Thaller's *kleio* is probably the best-known example of a DH approach to this: a database able to record features of the source that help to document data capturing decisions of the scholar.

Historical method respects the philological sincerity when handling text, but it aims for the information one can extract from the text. This leads to a specific form of digital scholarly editions, I called "assertive" edition (Vogeler 2019). However, the preference of historians for information reduces the interest in text as sequence of codepoints (or less technical: the sequence of symbols representing a coherent linguistic item), if this is just additional work. I would like to argue that the combination of visual reproduction and structured data representation of the content conveyed is similar to a scholarly edition, but not a scholarly edition in the proper sense - i.e. a "proto-edition" (Vogeler 2022).

The proto-edition is part of a pragmatic discourse on publishing in particular historical records: medievalists realised that the prac-

tice of deep engagement with a text is too time consuming to use it as a method for the publication of the increasing number of records preserved in the European archives since the 14th century. To address this "problem of the mass", Ivan Hlavacek (2006), Olivier Guyotjeannin (2005), and Benoit Tock (2008) suggested to publish (digital) images with deep descriptions of the content, e.g. following the practice of calendaring, a practice established in the medievalist community already in the 19th century to inform historians about the main "facts" reported in the records. These abstracts, digests, *regesta* are sufficient for most of the historians to build their arguments on. However, historians had to rely on the reputation of the people creating the calendars and could not dive deeper. Digital reproduction changes this significantly as adding an image of the record to the abstract is a minor economic effort. The digital surrogate empowers the experts to judge on their own.

The proto-edition shares with the critical scholarly edition its rigor: when the scholarly editors work on creating a completely consistent and reliable transcription, following best practices and discussion every minor philological detail, the "proto-editors" are engaged in the reliable and consistent representation of the content of a document. They try to use fixed terminology of re-occurring activities, pay attention to the identification of persons and places mentioned, are careful about alternative interpretations, and refer to scholarly literature discussing the document. They document their practice. In the digital world, they pay attention to re-usable data models, e.g. referring to upper level ontologies like the CIDOC-CRM, or incorporating the ideas of the Records in Context model. They don't rely on the archival description only, but they consider the document as a historical source in the research discourse to which the proto-edition is considered a contribution.

Calling this an edition (and even a "Proto-edition") raises the question: An edition without text, how can that be? There is a conceptual answer to this: With Patrick Sahle's pluralistic view (Sahle 2013), text can be perceived as a visual item. Or less elaborate: a photograph of manuscript, a book, or an inscription is certainly a text. Europeana has categorized the digital surrogates as TEXT since long. The visual item shares many features of other representations of text: One can transform it into other forms, e.g. by reading it aloud, can understand it, and can reflect on its form (as rhythm and rhyme, as narrative etc.). The proto-edition seems to be a topic for historians only. However, already the layout has semantics as it is reflected in the SegmOnto proposal by Gabay et al. (2021). Christoph Flüeler (2015) connected manuscript descriptions to scholarly editions by the depth of the description.

Finally, the proto-edition has to make its case in relationship to modern machine learning based methods to handle archival records in large numbers. Dominique Stutzmann has pointed out correctly at the IMC 2019 that for pure HTR and NLP based representations of texts, the term "artificial edition" would be a good fit: The representation created by the machine is not comparable to any kind of human intervention. Scheithauer et al. (2021) have already proposed encoding practices to record this. Even if accuracy is high - in terms of the previous evaluation of the methods - it is still carries a different bias than any human representation: While the machine uses models based on pre-existing human interpretations and unsupervised representations of other corpora, the human selection and interpretation represents the research interest and the creativity of the scholar, being part of a scholarly community. The proto-edition shares, and that is important, with the scholarly edition the human interaction with the historical texts.

## Bibliography

**Dillen, Wout** (2019). „On Edited Archives and Archived Editions“, in: *International Journal of Digital Humanities* 1, Nr. 2: 263–77. DOI: 10.1007/s42803-019-00018-4.

**Eggert, Paul** (2017). „The archival impulse and the editorial impulse“. In: *Advances in digital scholarly editing. Papers presented at the DiXiT conferences in the Hague, Cologne, and Antwerp*, ed. by Peter Boot, Anna Cappellotto, Wout Dillen, Franz Fischer, Adhoan Kelly, Alexander Mertgens, Anna-Maria Sichani, Elena Spadini, and Dirk Van Hulle. Leiden: Sidestone Press.: 121–24.

**Flüeler, Christoph**. (2015) “Digital Manuscripts as Critical Edition”, in: *The Schoenberg Institute for Manuscript Studies - Blog*. <http://schoenberginstitute.org/2015/06/30/digital-manuscripts-as-critical-edition/>.

**Gabay, Simon / Camps, Jean-Baptiste / Pinche, Ariane / Jahan, Claire** (2021). SegmOnto: common vocabulary and practices for analysing the layout of manuscripts (and more). *1st International Workshop on Computational Paleography (IWCP@ICDAR 2021)*, Lausanne, Switzerland. (hal-03336528)

**Guyotjeannin, Olivier** (2005). „Éditions diplomatiques et recherche historique: Quelques remarques sur le cas français (XIXe–XXe siècles)“, in: *Vom Nutzen des Edierens*, ed. by Brigitte Merta, Andrea Sommerlechner, und Herwig Weigl. MIOG Erg. Bd.47. Wien: Oldenbourg:303–12.

**Hlaváček, Ivan** (2006). „Das Problem der Masse: Das Spätmittelalter“, in: *Archiv für Diplomatik, Schriftgeschichte, Siegel- und Wappenkunde* 52: 371–93.

**Kaplan, Frédéric / Isabella di Leonardo** (2017). „Big Data of the Past“, in: *Frontiers in Digital Humanities* 4. DOI: 10.3389/fdigh.2017.00012.

**Pédaque, Roger T.** (collective) (2007). *La redocumentarisation du monde*. Toulouse: Cépaduès.

**Price, Kenneth M.** (2009). „Edition, Project, Database, Archive, Thematic Research Collection: What’s in a Name?“, in: *Digital Humanities Quarterly* 3, Nr. 3. <http://digitalhumanities.org:8081/dhq/vol/3/3/000053/000053.html>.

**Sahle, Patrick** (2013). *Digitale Editionsformen: zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels. Teil 3: Textbegriffe und Recodierung*. Schriften des Instituts für Dokumentologie und Editorik 9. Norderstedt: BoD, Books on Demand.

**Sahle, Patrick** (2007). „Digitales Archiv - Digitale Edition. Anmerkungen zur Begriffsklärung“, in: *Literatur und Literaturwissenschaft auf dem Weg zu den neuen Medien. Eine Standortbestimmung*, ed. by Michael Stolz. Norderstedt: Books on Demand: 64–84.

**Scheithauer, Hugo / Chagué, Alix / Romary, Laurent** (2021). “From eScriptorium to TEI Publisher. Brace your digital scholarly edition!”, Berlin, France. ( hal-03538115 )

**Tock, Benoît-Michel** (2008). „Rev. of: B. Merta, A. Sommerlechner, H. Weigl (Hg.): Vom Nutzen des Edierens“, in: *Francia Recensio*, Nr. 1. [https://perspectivia.net/receive/ploneimport2\\_mods\\_00007748](https://perspectivia.net/receive/ploneimport2_mods_00007748)

**Van Hulle, Dirk** (2009). „Editie en/of Archief: modern manuscripten in een digitale architectuur“, in: Verslagen en Mededelingen van de Koninklijke Academie voor Nederlandse Taal- en Letterkunde 119, Nr. 2: 163–78.

**Vogeler, Georg** (2019). „The ‘Assertive Edition’“. *International Journal of Digital Humanities* 1, Nr. 2: 309–22. DOI: 10.1007/s42803-019-00025-5.

**Vogeler, Georg** (2022). „Edition – Protoedition – Reproduktion : Der digitale Wandel“, in: *Geschichte in Wissenschaft und Unterricht* 73, Nr. 9/10: 498–511.