

== DRAFT ==

The ‘assertive edition’¹

On the consequences of digital methods in scholarly editing for historians

Georg Vogeler, georg.vogeler@uni-graz.at

forthcoming in International Journal of Digital Humanities 1 (2018)

Abstract

The paper describes the special interest in edition by historians and the resulting editorial practice in contrast to the methods applied by pure philological textual criticism. The interest in historical ‘facts’ suggests methods trying to create formal representations of the information conveyed by the text in structured databases. This can be achieved with RDF representations of statements extracted from the text, by automatic information extraction methods or by hand. The paper suggests the use of embedded RDF representations in TEI mark-up, following the practice in several recent projects and concludes with a proposal for a definition of the ‘assertive edition’.

Introduction

Historians have a different approach to scholarly editing than literary scholars. Both share an interest in a good text created by textual criticism, as the main source of historical intelligence comes from texts. Nevertheless, historians can have a slightly different approach to text: linguistic and physical aspects are considered mere intermediates to the information conveyed by the text. Historians consider the content of the text as ‘data’, and they want to use this data in their research to gain knowledge about the past. The circumstances under which archival documentation as a major type of text historians are working with, is created, supports their perception of text: people recorded administrative activities in text to preserve information on these activities for contemporary but absent or for future clerks – they stored their data in texts written on paper.

In pre-digital editorial practice this can lead to decisions literary scholars could not accept, like paraphrasing parts of the text. I will try to show that in digital scholarly editing the historians approach to editing can be reconciled with methods of textual scholarship. I suggest calling this combined method ‘assertive edition’ to avoid the impression that this method is only applicable by historians. The method of assertive editing is not defined by disciplinary interests but by an interest in one facet of text: the information recorded. In terms of Patrick Sahle’s text wheel (Sahle 2013:III,45-49) the assertive edition is the editorial practice dedicated to the ‘text as content’ perspective. It has to be noted that in the following I usually will oppose this ‘content’ to the ‘text’ as pure transcriptions and the result of text critical work.

Contributions to the assertive edition

¹ This text profited very much from the suggestions of the reviewers, the colleagues from the Institut für Dokumentologie und Editorik, and in particular of the hard work of Sean Winslow, for which I am deeply grateful.

Assertive editing is fed by two streams in pre-digital and early digital scholarship. The pre-digital editorial practice in historical research contributes the ideas of content-oriented navigation, the possibility of multiple forms of representation, and extensive historical commentary. I will try to show this by presenting on three major German historical printed editorial series: The Monumenta Germaniae Historica (MGH), the Records of the Early Modern Imperial Diet (“Reichstagsakten”), and the Official Minutes from the Imperial Chancellery (“Akten der Reichskanzlei”: Bundesarchiv 1982).

Pre-digital contributions

To facilitate navigation and reception the editions in the MGH Diplomata series prepends abstracts of the legal core to each document. This is common practice in European charter editions and codified by a committee of historical editors under the direction of Robert-Henri Bautier in 1974 (Bautier 1976:13, 17). More recent Diplomata series editions dedicate a paragraph in the introduction of each document to the historical context (e.g. the charters of Emperor Frederic II by Walter Koch, 2002-2017). Some MGH editions of historiographical texts indicate the year the current text is referring to in the margins (e.g. Georg Waitz’ edition of the *Historia Danorum Roskildensi*, 1892:21-26). This helps the readers to find the events, which they are interested in.

Of course, abstracts serve more purposes than simple navigation. In the editions of the Imperial Diets, abstracts replace some of the documents (e.g. Heil, 2014, 87-91). Dietmar Heil describes the interest of the editors: “The priority is ... philological authenticity, but optimal accessibility” (2015, 29, trans. Georg Vogeler) They reduce historical orthography and changes punctuation when it deviates from the modern syntactical analysis of the text. (Heil, 2015, 29-31). Editors of correspondence have also considered this approach (Steinecke, 1982).

Editorial work in contemporary history is defined by the selection of significant material and contextualization of the text. The editors of the Minutes of the Cabinet of the German Federal Government (Bundesarchiv, 1982), for instance, explain their selection by “inhaltliche Relevanz” (“if relevant”), discarding as irrelevant for instance the agenda in the head of each minute, invitations, and their attachments. This content-oriented approach can be found in other editorial principles of this edition: Orthographic and syntactic errors are emended without notice. Single entries start with a heading, persons present, and the place and time of the meeting, not as a verbal copy but as an extract created by the editors.

The Minutes also serve as an example for the third element of pre-digital editorial practice: They add extensive notes on the subjects of the meetings to each transcript with the aim of making the texts understandable. This kind of annotation is not specific to this one edition, but is generally recommended in historical editing (Cullen, 1981; Stevens/Burg, 1997, 157). The edition of the Minutes of the Bundeskabinett serves to illuminate government decisions, rather than their wording. Similarly, many MGH editors add extensive comments on the historical context, e.g. in the pre-publication of anonymous continuations of Frutolf 1101-1106 (Marxreiter, 2018).

These approaches have been directly transferred into electronic editions. The idea of facilitating the understanding of the text accepts translations as way of edition. This leads to solutions like David Postles online representation of Stubbington medieval records (Postles, 2011), which gives the text in a translation of the original Latin text. This not an individual practice, as P.D.A. Harvey discusses in his introduction into historical editing translation as method in editions too (2001, 31-32). From a historians point of view a translation is a sensible solution as it facilitates the use of the document. It would not satisfy the research interests of textual scholars.

Paul D.A. Harvey argues that the edition of historical records can be reduced to a calendar of abstracts, when the original or photocopies of them are easily accessible (2001, 56-59). Several

project follow an approach of this kind: The Records of the Swiss Foreign Office (Zala et al., 1979-2018) replaces the transcription with images. This calendar plus image approach is also used by Soundtoll registers (Veluwenkamp / Woude, 2009; Gøbel, 2010), and Peter Rauscher and his colleagues in the *Donauhandel* project (2008-2018): Both create databases with structured information directly from the source and link it to images of the source.

Early digital contributions

Historians interests in the ‘facts’ and the dominance of sociologic approaches to history in the 1960s to 1980s led them to create ‘databases’ of historical information (Boonstra et al., 2004). A famous example for this approach is the Online Catasto of 1427 (Herlihy et al., 2002), an online edition created by R. Burr Litchfield and Anthony Molho, based upon David Herlihy and Christiane Klapisch-Zuber’s project *Census and Property Survey of Florentine Dominions in the Province of Tuscany, 1427-1480* (1978; Herlihy, 1964; Herlihy, 1967). The data keeps close to the source copying the information on wealth recorded for each taxable household in the city (as they can be found in the initial tax declarations of 1427 plus additions and adjustments made in 1428 and 1429). Seeing historical records as an accidental medial solution to preserve and process information, one could consider this database as a simple change in recording medium not in information itself. The needs of the recording medium require substantial changes in the recording method: Herlihy / Klapisch-Zuber had to create new encodings and had to break the text rigorously into table columns. In the end, the database tries to recreate the information recorded by the Florentine officials: *who* had to pay *what* amount of taxes for *which* kind of property.

Philological editors certainly cannot consider this database an edition. The encoders did not copy family names, names, and patronymics letter by letter, but standardized them, and truncated them beyond ten letters. Historians were well aware of the modifications that database encoding made to the original records. In the 1970s, however, digital scholarly editing was not yet developed enough to provide a solution. The concept of scholarly editing does not even appear in the more recent book on Historical Information Science by Lawrence McCrank’s (2002). At the time, computing methods in the historical sciences chiefly meant the production of relational databases and spreadsheets.

In the 1980s, Manfred Thaller proposed a historical database system that kept closer to the original source (1980; 1988; 1992; 1993). He developed the *Clio* database as a ‘source oriented’ database. It would reduce the amount of encoding and transformation of the source customary. *Clio* kept as much information from the source as possible by allowing for hierarchical organization of information, better representation of incomplete data, and integration of alternatives and comments. This source oriented database approach is clearly a type of editorial work, combining text from the source with interpretation for historians. At the same time, a philologist would regret the lack of a full transcription.

Digital Editions and Facts

Digital Scholarly Editing has developed since the days of *clio* and built upon the methods developed for the MGH, the Reichstagsakten, and the Akten der Reichskanzlei. The assertive edition developing out of these strands is something between pure textual representations and well-formed databases built around specific research questions. No edition yet calls itself an assertive edition, but many carry features that fit the definition put forward here. A selection may be found by searching Patrick Sahle’s catalogue for “general subject area: history” (Sahle. 2008-2017). Browsing through the projects in the list, one can identify four major questions:

1. Which interface elements are typical for an assertive edition?

2. How can we use automatic information extraction processes into the scholarly edition?
3. Is semantic mark-up (provided by the TEI) sufficient?
4. How can we integrate the Web of Data (the 'Semantic Web') into scholarly editions?

Interface elements

Editions like the letters of Alfred Escher (Jung, 2012-2018), the Acta Pacis Westfalica (Lanzinner / Braun, 2014), and the Diplomatic Correspondence of Thomas Bodley 1585-1597 (Adams, 2011), offer avenues of access to the text beyond the pre-existing textual-structure. Typically, tools include indices of persons, places, and subject keywords. Other entry points to the texts show better what an assertive scholarly edition could be concerned with: APW, for instance, gives access via a timeline of events, a calendar of relevant dates, and a map. Indeed indices of persons, places, and events, calendars and maps are fast becoming default components for historical digital editions. Additional fact-oriented interface elements seem to depend more on the type of documents edited: rich prosopographical information like in correspondence suggests using network visualisations as in the diplomatic correspondence of Thomas Bodley (Adams, 2011, visualisations). Economic information suggest the use of bar charts to visualize income and expenditure, as in the edition of the municipal accounts of Basel 1535-1611 (Burghartz, 2016, Konten). The latter builds upon the source-oriented database approach advocated by Manfred Thaller by allowing the user to select entries from the accounts and collect them in a 'data basket' (Burghartz, 2016, databasket). This allows the user to perform basic arithmetic operations, and download as a spreadsheet. Finally, semantic networks like those used in Burkhardt Source (Ghelardi et al., 2015) hold a general promise, but currently remain lonely solutions for single projects.

Information extraction

The user interfaces is, of course, only the surface of the edition. How does one harvest information? What form does the information take as digital data? Which models relate the information to the transcription? One approach to data harvesting from texts is automatic information extraction. Computer linguists have been working on this since the 1950s. Their goal is to reduce free prose text to answers to the question "*Who did what to whom when?*" and represent these answers in a structured way. A typical information extraction pipeline starts with generic Natural Language Processing steps, and then uses Named Entity Recognition to mark-up the words representing persons, locations, or organizations, temporal data, and quantifying data. The pipeline then relates these entities to each other, building connections between the entities. This can take the form of predicates in sentences, co-references by pronouns, etc. The possible relationships can be inferred from external knowledge about the domain, like dates of birth and death for people mentioned in a text, or it can be the result of the semantic role, such as can be inferred from the predicate in a sentence. The task is very domain-specific, as it depends on what type of information is considered relevant. A typical task for historical research could be event extraction, which is already applied to automatic news analysis (see Grisham, 2015 for a general introduction).

Recent projects dealing with US foreign affairs records have taken this approach to transcripts of archival documents. They take the historical records as source data without any intermediate scholarly processing. Using OCR to create a digital representation of the text, scholars then apply distant reading methods like topic modelling or information extraction to this corpus (e.g. Kaufmann, 2014-2018). Gao et al (2017) have used even used the electronic texts of the cables in the 1970s for their computer based analysis.

Implementing this approach in scholarly editions, the aim would be to create a reliable text with classical textual criticism and extract from this text the information for historians. Existing information extraction methods are built for modern texts, and thus have to be modified to be

applicable to historical texts or historical texts have to be modified to come closer to modern texts. Michael Piotrowsky (2012) has described the many challenges in this task. Some progress has been made e.g. in the handling of variants in historical language, as by Bryan Jurish (2008; 2010; 2011; 2013), or Kestemont et al. (2017). However, most of the problems are still to be solved. Scholarly editors still have to rely on their own competence and on human labor for the introduction of substantial knowledge about what people in the past wrote in their texts.

TEI and Semantic mark-up

The problems computers still have with historical languages led to the decision to create manually annotated texts. Digital editions use the extensible mark-up language XML to add semantic mark-up to texts. This is in particular made possible by the strong connection between the communities maintaining the guidelines of the Text Encoding Initiative (TEI) with the community of digital scholarly editors. TEI provides semantic annotation for many phenomena interesting to historians: names of persons, locations, or organizations can be encoded as `<name>`, temporal expressions as `<date>` and `<time>`. With the TEI P5 there are even guidelines how to encode structured descriptions of persons, places, and events, structures that are similar to database structures. Still, the mark-up provided by the TEI is deficient in ways to express historical information of interest in this present study. An example is the `<event>` element.² The TEI guidelines consider it as a concept independent from text, to which text can refer. An expression like “my inauguration” in “after my inauguration, I decided to leave the town” is not an ‘event’ but should be encoded like any other referring string with the `<rs>` Element. Nevertheless, while places and persons have a dedicated `<persName>/<placeName>` tagging, historians interested in marking up named events like “World War I”, the “battle of Marathon”, the “coronation of Charlemagne”, the “Contract of Maastricht”, or the “Lisbon Earthquake” in their sources have to employ workarounds. This observation illuminates the distance between a major practice in digital scholarly editing and research interests of historians. One reason for this might be, that the scholars much more easily agree on the identification of individual names of concrete persons, places, and organizations than on more abstract events. The sample events above have formal names (some, more than one), but text often describes events in a much looser way: “my inauguration as bishop” is clearly an individual event, but one unlikely to have formalized name. Many events even do not carry names at all, but they are told as a story: “When Hitler’s troops crossed the Polish border on September 1 in the year 1939, World War II had started.” This sentence clearly refers to the event “Nazi invasion of Poland”, but could just as easily be referred to as “Start of World War II” or in many other ways. This example demonstrates that even these short identifiers are not just an arbitrary ‘name’. They create different contexts and are therefore part of a specific discourse.

Web of Data: Semantic Mark-up by reference

Linking different names for the same event are a typical competency of Semantic Web technologies as proposed by the W3C since 2001 (Berners-Lee / Hendler / Lassila, 2001; rebranded as “Web of Data” activities by W3C, 2013). The Semantic Web uses abstract unique identifiers (URIs) as representations of the concepts covered by the name. With URIs, scholars can create digital representations of events without relying on ambiguous natural language terms. An increasing number of digital scholarly editions use Semantic Web technologies to solve naming issues. The most prominent method is the extension of classical indices: while previously, such indices standardized names to represent the historical fact behind a name for a person or a place, URIs allow identifying persons, places and organisations for technical processing, even if there is no name. Gautier Poupeau described this approach in 2006, and the digital edition of the Fine Rolls of King Henry III, created 2005–2011, made extensive use of these technologies in its back-end. A good example of the use of

² <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/ref-event.html>

Semantic Web technologies in scholarly editions is the *Teutsche Academie der Bau-, Bild- und Mahlerey-Künste*, by Joachim von Sandrart (Kirchner et al., 2008–2012). The text refers to many artists and artistic objects, which are identified and described in the index, and can be downloaded from the site as an RDF dataset³.

A more extensive formalisation than the index approach is demonstrated by the *Old Bailey* project (Hitchcock et al., 2003-2018). The basic transcription of the text was annotated in XML in order to facilitate structured searching and statistical analysis. This approach works because the records already tend to have a regular structure. 'The meaning of particular words or phrases like names and crimes is tagged, and further sorted into subcategories like types of verdict.'⁴ The final encoding of the texts contains formal descriptions of the relationships established by the mark-up. They are processed in a separate database, but kept together with the text in the XML. *Old Bailey online* is not just a database of criminal trials, but an assertive scholarly edition, representing the statements made by the transcription in a formal way and linking the statements to transcription, to image.

Following Semantic Web / Web of Data activities of the W3C, the digital representation of data is increasingly realized through RDF triples. In the context of the assertive edition they have the advantage that they model facts as statements about the reality in a simple but expressive way, as they can be read as *subject predicate object* propositions.

Parallel to the development of embedded annotation with XML digital humanities has developed methods for stand-off-annotation. Since 2001, they are increasingly realized with RDF. They have found a standard in the Open Annotation vocabulary (Sanderson/ Ciccarese / Van de Sompel, 2013). Digital editions have made use of this possibility: Pund-it (Morbidoni, 2015; Net7) is the most advanced application of the Semantic Web to digital scholarly editions, used for example in the scholarly edition of the correspondence of Jakob Burckhard (Ghelardi et al., 2015). It allows annotation of any part of the text. Textual fragments may be used as the subjects or objects of an RDF triple. In the Jakob Burckhard edition, Pund-it reduces the possible predicates to references to artworks and artists, general comments, quotations and references, dates, and geographical identification. Pundit saves this as an RDF reference to the HTML elements. Work is under way on linking the annotation directly to the XML/TEI source.⁵ In the end, the semantic networks, which are a unique interaction feature of this digital edition, can describe the content of the text through direct links to part of the source text that contains the information.

How to combine transcription with databases?

Looking forward, a number of questions arise: can we build scholarly editions which include results similar to those created by information extraction software but controlled by hand, thus bringing the full power of human understanding to the annotation? Can we encode the propositions made by the words of a text into the transcription? Can we embed the statements extracted by the reader into the sequence of characters and thus create a single digital resource representing transcription and information conveyed by it to the editor? If so, how?

One possible approach is suggested by RDFa, the W3C's proposed serialization of RDF embedded in HTML mark-up. It provides attributes for HTML elements describing RDF triples. Existing HTML element attributes like @href or @src can be used as objects in the 'subject predicate object' triple

³ <http://ta.sandrart.net/data/> and single expressions via the REST API of the project at <http://ta.sandrart.net/de/info/services/rest/>, so <http://ta.sandrart.net/services/rdf/person/561> returns the RDF data for Philipp Melanchthon for example.

⁴ <https://www.oldbaileyonline.org/static/Project.jsp#mark-up>

⁵ <https://net7.github.io/pundit2/xmltei.html>

structure. Additional attributes like @typeOf, @resource and @property permit some of the full expressiveness of RDF.

```
<p xmlns="http://www.w3.org/1999/xhtml">
  <span resource="GND:Gleim">Bertrand marched out with the <span
  property="ex:isPartOf" resource="ex:Guards">Guards battalion to <span
  property="ex:marchTo" resource="geo:Wrocław">Breslau</span></span> and writes
  <span property="ex:writes" resource="_:Briefe"><span
  property="ex:emotionalQuality" resource="ex:Serene">gay</span>
  letters</span></span></p>
```

Listing 1a: example of a sentence in RDFa encoding

```
GND:Bertrand ex:isPartOf ex:Guards ;
  ex:marchTo geo:Wrocław ;
  ex:writes [ ex:emotionalQuality ex:Serene ] .
```

Listing 1b: triples extracted from the sentence (in Turtle/N3 notation)

Listing 1 demonstrates which triples can be extracted from a sentence in a fictive letter by using RDFa mark-up as semantic annotation. This method is attractive, as it closely relates the assertive expression to the text.

How might TEI be similarly extended? The standardized mark-up of the TEI covers some typical basic facts that might be extracted from texts. However, assertive annotation can be much richer and highly diverse. Something more flexible is needed. Therefore, I would suggest transferring the RDFa approach to TEI, creating a 'TEIa' annotation style. The TEI-community discussed already the idea of directly importing the RDFa attributes into the TEI (TEI-Community, 2010; 2014), but it was argued convincingly that foreign namespace not controllable by TEI and therefore not recommended. Fortunately TEI provides attributes, which cover much of a TEIa approach: @ref creates a link from a verbal expression to an entity and @ana links textual fragments to any kind of analytical annotation. As the TEI guidelines reduce the use of @ref to reference strings, the globally usable @ana seems to be the best candidate for a generic linking of textual fragments to RDF triple structures describing the relevant facts.

The *Système Modulaire de la Gestion d'information Historique* (SyMoGIH), developed by Francesco Beretta and his team (Beretta, 2012; Beretta et al., 2016), makes use of RDF based semantic mark-up in combination with TEI transcriptions⁶. In the edition of the Journal of Léonard Michon (Lericot, 2017), for instance, the transcribed texts are accompanied by a marginal index with short notes on events, facts and persons. They are formalisations of the very text, e.g. « Le Roy luy a envoyé à Marseille Monsieur de Saint Olon, gentilhomme ordinaire, qui le suivra jusqu'à Paris, »⁷ is represented by a descriptive text « François Pidou de Saint Olon accompagne l'ambassadeur perse de Marseille à Versailles. » and the people involved in the event. This information is represented as an RDF statement (<http://symogih.org/resource/Info116905>) about the two persons involved. The annotation is encoded with the TEI and the global attribute @ana links the text to a database of the formalized description of the content (Beretta, 2015).

Other examples for this approach can be found in projects realized at the Zentrum für Informationsmodellierung at University of Graz in cooperation with the Historical Department at the University of Basel. Susanna Burghartz' team created transcriptions of two sets of administrative records from the city of Basel from the Early Modern Period: the annual accounts of the city from

⁶ <http://symogih.org/>

⁷ http://journal-michon.symogih.org/documents/document-text.html?id=DiOb5072&volume_number=1&page_number=270

1535 to 1611 (Burghartz, 2016) and a criminal court record, the 'Urfehdebuch' (register of oath of truce) from 1563-1569 (Burghartz / Calvi / Vogeler, 2017). While Digital Humanities projects related to the early modern period focus very often on handling the specific properties of early modern texts (Neslon / Terras, 2012; Estill et al., 2016), the Basel editions can be considered assertive editions. Both projects are realized in a very flexible technical environment, the GAMS (Steiner / Stigler, 2014-2017), which is a framework for archiving and publication of humanities data sources, in particular digital scholarly editions.⁸ In the *Jahrrechnung der Stadt Basel* the core information unit addressed is clear: the monetary amount of a single transaction, as transmitted by the historical accountant, i.e. his rubrics (Vogeler, 2015a; Vogeler 2015b; cfr. Vogeler, 2016 for a deeper discussion of editorial methods appropriate for historical accounts). However, even this simple criterion needs interpretation: repaid loans and interests are mixed in one common category of income. For a financial analysis this is unacceptable; accordingly, stand-off annotation is used to apply sub-categories to individual entries. In the case of the *Urfehdebuch* the main category is – similar to the *Old Bailey* records – a single case. At least one core property of the data structure is already represented by the textual structure in the archival manuscript: the heading gives the name of the offender. However, type of offence, victim, and punishment have to be extracted from the text and are encoded with links to a taxonomy developed for the project (Pollin / Vogeler, 2017).

Embedding the interpretation of facts into the text seems straight forward, but has several drawbacks. The examples above have shown that we need at least links to external knowledge organization systems and the translation into full RDF triples to express enough of the content of the texts. Information science teaches us to go even further and to include time in the relationship between data and information, i.e. between edited text and the facts the historian considers to be represented by the text. Börje Langefors (1966/1973) has formulated his 'infological equation', in which the information is a function of the data, the recovering structure, and of the time when the interpretation takes place. This conceptualization of information argues in favor of stand-off annotation as the semantic value of an edited text is an interpretation by the editor. In fact, there is a long-standing discussion in the text encoding community on the risks of embedded semantic markup, summarized by Manfred Thaller and Dino Buzzetti (2012). Standardized technical solutions for this approach do not exist yet. RDF has established itself as a common data structure for the exchange of the factual interpretations of text. How to maintain the linkage between the text edited and the RDF is still under discussion.

Conclusion

All of this leads to the following description of assertive editions: they are scholarly representations of historical documents in which the information on facts asserted by the transcription is in the focus of editorial work. They help the user/reader to understand the text and to use the information conveyed in the text as structured data. This data includes interpretations of the text drawn from context and expertise of the editor. In fact, interpretation is in the core of the critical activity of the editor: She concludes from her knowledge about the written text, its layout, and historical circumstances, under which it was produced, how to describe the content beyond pure transcription. This can include normalization, categorization, reference to external resources, formal knowledge representations, and many other forms of transformation.

The assertive edition is not a well-defined type of scholarly editing yet. However, assertive editions exist. The methods to create them, to model them, and to make them available online are becoming

⁸ <http://gams.uni-graz.at>

a part of scholarship. Digital assertive editions can be identified by the user interface as well as in the data structures, which try to combine the transcription with a database of statements made in the text. On the one hand, few historians have already implemented the concept: It allows them to employ source-oriented critical methods while working with large amounts of data. The majority of the historians still focus on the structured data extracted from the sources. Databases are their major tool, often employing rich interfaces and elaborate visualizations. The majority of scholarly editors on the other hand employ traditional methods of textual scholarship; they ponder complex transcription problems, evaluate variants, and include textual materiality. The combination – deep links between structured data and text with assertive editing – is still rare. One reason for this is the technological ability to realize them. Tools like Pund-it, frameworks like Symogih or GAMS, and best practice examples like the projects cited above, are about changing this.

References⁹

- Adams, R.** (ed.) (2011). *The Diplomatic Correspondence of Thomas Bodley*. Centre for Editing Lives and Letters. Retrieved from <http://www.livesandletters.ac.uk/bodley/bodley.html>
- Andreini, G., Di Donato, F., Giacomi, D., Giusti, E., & Masotti, R.** (2016). Pundit. Semantic Annotation for Digital Humanities. *Digital Humanities 2016: Conference Abstracts*. Jagiellonian University & Pedagogical University, Kraków, pp. 728-729. Retrieved from <http://dh2016.adho.org/abstracts/363>
- Association for Documentary Editing (ADE)** (s.d.). About Documentary Editing. Retrieved from http://www.documentaryediting.org/wordpress/?page_id=482
- Beretta, F.** (2013). The symogih.org project and TEI : encoding structured historical data in XML texts. *Text Encoding Initiative Conference and Members' Meeting 2015. Connect, Animate, Innovate.*, Oct 2015, Lyon, France. Retrieved from <https://halshs.archives-ouvertes.fr/halshs-01251915>
- Beretta, F., Butez, C. C., Carpentier, A., & Delcourte, Marie** (2016). Reconstituer les évolutions des espaces forestiers de l'Avesnois aux XIVe – XVIIIe siècles. Approches méthodologiques. *Bulletin du centre d'études médiévales d'Auxerre | BUCEMA*, Hors-série 9, Retrieved from <http://cem.revues.org/13774>
- Beretta, F., & Vernus, P.** (2012). Le projet SyMoGIH et la modélisation de l'information: une opération scientifique au service de l'histoire. *Les Carnets du LARHRA* 1, 81-107.
- Berners-lee, T., Hendler, J., & Lassila, O.** (2001). The SemanticWeb. *The Scientific American* 284, 5, 34–43.
- Boonstra, O., Breure, L., & Doorn, P.** (2004). *Past, Present and Future of Historical Information Science*, Amsterdam: DANS. Retrieved from <http://www.oopen.org/search?identifiant=353255>
- Bundesarchiv** (1982 ff.). *Die Kabinettsprotokolle der Bundesregierung*, München. Retrieved from <http://www.bundesarchiv.de/cocoon/barch/0000/index.html>
- Burghartz, S.**(ed.) (2015). *Jahrrechnungen der Stadt Basel 1535-1611 – digital*. Basel/Graz. Retrieved from <http://hdl.handle.net/11471/1010.1>
- Burghartz, S., Calvi, S., & Vogeler, G.** (ed.) (2017). *Urfehdebücher der Stadt Basel - digitale Edition*. Basel/Graz. Retrieved from <http://hdl.handle.net/11471/1010.2>
- Ciccarese, P., Ocana, M., & Clark, T.** (2012). DOMEQ: a web-based tool for semantic annotation of online documents. *Bio-Ontologies* 2011.

⁹ All URLs were last checked on Apr. 2, 2018 and if possible archive via the web archiving service of <http://archive.org>.

- Ciula, A., Spence, P., & Veira, J. M.** (2008). Expressing complex associations in medieval historical documents. The Henry III Fine Rolls Project. *Literary and Linguistic Computing* 23, 311-325. Retrieved from <http://www.finerollshenry3.org.uk/>
- Cullen, C. T.** (1981). Principles of Annotation in Editing Historical Documents; or, How to Avoid Breaking the Butterfly on the Wheel of Scholarship. *Literary and Historical Editing*, ed. by Vogt, G. L., & Bush Jones, J. Kansas: Univ. of Kansas Libraries (University of Kansas Publications Library Series, 46).
- Estill, L., Jakacki, D. K., & Ulliyot, M.** (2016). *Early Modern Studies after the Digital Turn*, Toronto.
- Gao, Y., Goetz, J., Mazumder, R., & Connelly, M.** (2017). *Mining Events with Declassified Diplomatic Documents*, 20 Dec 2017. Retrieved from <https://arxiv.org/abs/1712.07319>
- Gerber, A., & Hunter, J.** (2010). Authoring, editing and visualizing compound objects for literary scholarship. *Journal of Digital Information*, [S.l.], v. 11, n. 1, mar. 2010. Retrieved from <https://journals.tdl.org/jodi/index.php/jodi/article/view/755>
- Ghelardi, M.** et al. (2015). *Burkhardt Source*. <http://burckhardtsource.org/>
- Gøbel, E.** (2010). The Sound Toll Registers Online Project, 1497-1857. *International Journal of Maritime History* XXII, 2, p. 305-324.
- Grassi, M., Morbidoni, C., Nucci, M., Fonda, S., & Piazza F.** (2013). Pundit: Augmenting web contents with semantics. *Literary and Linguistic Computing*. 28. 640–659. DOI: 10.1093/lc/fqt060
- Greetham, D.** (2008). What is Textual Scholarship? *A Companion to the History of the Book*. Ed. by Eliot, S., & Rose, J. 21-32. DOI: [10.1002/9780470690949.ch2](https://doi.org/10.1002/9780470690949.ch2)
- Grishman, R.** (2015). Information Extraction. *The Oxford Handbook of Computational Linguistics*. Ed. by Mitkov, R. 2nd edition. Oxford University Press. DOI: 10.1093/oxfordhb/9780199573691.013.009
- Heil, D.** (ed.) (2014). *Der Reichstag zu Konstanz 1507*. München: Historische Kommission 2014 (Deutsche Reichstagsakten / Mittlere Reihe / Deutsche Reichstagsakten unter Maximilian I. 9). Retrieved from <http://reichstagsakten.de/index.php?vol=rta1507>
- Heil, D.** (2015). Per aspera ad acta. Ein Werkstattbericht zur Edition der Deutschen Reichstagsakten aus der Zeit Kaiser Maximilians I. *Nit wenig verwunderns und nachdenkens. Die 'Reichstagsakten - Mittlere Reihe' in Edition und Forschung*. Ed. by Wolgast, E., Göttingen (Schriftenreihe der Historischen Kommission bei der Bayerischen Akademie der Wissenschaften 92), p. 19-43. DOI: 10.13109/9783666360831.19
- Herlihy, D.** (1964). Direct and indirect taxation in Tuscan urban finance, ca. 1200-1400. *Finances et comptabilité urbaines du XIII e au XVIe siècle*. Actes du Colloque International Blankenberge 6.-9. Sept. 1962, Brüssel: Pro Civitate (Historische Utigaven in-8 7), p. 385-405.
- Herlihy, D.** (1967). *Medieval and Renaissance Pistoia. The Social History of an Italian Town, 1200-1430*. New Haven: Yale Univ. Press.
- Herlihy, D., Klapisch-Zuber, C., Litchfield, R. B., & Molho, A.** (eds.) (2002). *Online Catasto of 1427*. Version 1.3. [Machine readable data file based on D. Herlihy and C. Klapisch-Zuber, *Census and Property Survey of Florentine Domains in the Province of Tuscany, 1427-1480*.] Florentine Renaissance Resources/STG: Brown University, Providence, R.I., 2002. Retrieved from <http://cds.library.brown.edu/projects/catasto/>
- Hitchcock, T., Shoemaker, R., Emsley, C., Howard, S., & McLaughlin, J.** (2003-2018) et al., *The Old Bailey Proceedings Online, 1674-1913*. Retrieved from <http://www.oldbaileyonline.org>
- Jung, J.** (ed.) (2012-2018). *Alfred Escher-Briefedition*. Retrieved from <https://www.briefedition.alfred-escher.ch/>
- Jurish, B.** (2008). Finding canonical forms for historical German text. *Text Resources and Lexical Knowledge*. Ed. by Storrer, Geyken, Siebert, & Würzner, Proceedings KONVENS. Berlin, p. 27-37.

- Jurish, B.** (2010). More Than Words. Using Token Context to Improve Canonicalization of Historical German. *Ldv Forum - LDV* 25, p.23-39. Retrieved from http://www.jlcl.org/2010_Heft1/bryan_jurish.pdf
- Jurish, B.** (2011). *Finite-state canonicalization techniques for historical German*. Potsdam. Retrieved from <http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-55789>
- Jurish, B.** (2013). Canonicalizing the Deutsches Textarchiv. Hafemann, Ingelore (Ed.): *Perspektiven einer corpusbasierten historischen Linguistik und Philologie*. Internationale Tagung des Akademienvorhabens 'Altägyptisches Wörterbuch' an der BBAW, 12.-13. Dezember 2011, Berlin (Thesaurus Linguae Aegyptiae 4), p. 235-244
- Kaufmann, M.** (2014-2018). "Everything on Paper Will Be Used Against Me": Quantifying Kissinger. Retrieved from <http://blog.quantifyingkissinger.com/>
- Kestemont, M., Pauw, G. de, Van Nie, R., & Daelemans, W.** (2017). Lemmatisation for Variation-Rich Languages Using Deep Learning. *Digital Scholarship in the Humanities*, Volume 32, Issue 4, 1 December 2017, p. 797–815. DOI: <https://doi.org/10.1093/lc/fqw034>
- Koch, W.** (ed.) (2002-2017). Die Urkunden Friedrichs II., currently 5 vols. Hannover. (MGH DD 14).
- Langefors, B.** (1966). Theoretical Analysis of Information Systems, Studentlitteratur, Auerbacher.
- Lanzinner, M., & Braun, G.** (eds.) (2014). Acta Pacis Westfalica digital. Retrieved from <http://apw.digitale-sammlungen.de/>
- Letricot, R.** (ed.) (2017). Édition critique numérique des Mémoires de Léonard Michon, *Université Lyon 3, LARHRA (CNRS UMR 5190)*. Retrieved from <http://journal-michon.symogih.org/index.html>
- McCrank, L. J.** (2002). Historical Information Science. An Emergin Undiscipline. Medford, NJ.
- Morbidoni, C., & Piccioli, A.** (2015). Curating a Document Collection via Crowdsourcing with Pundit 2.0. Gandon, F., Guéret, C., Villata, S., Breslin, J., Faron-Zucker, C., & Zimmermann, A. (eds). *The Semantic Web: ESWC 2015 Satellite Events. ESWC 2015. Lecture Notes in Computer Science*, vol 9341. Springer, Cham.
- Nelson, B. H., & Terras, M.** (eds.) (2012). Digitizing Medieval and Early Modern Material Culture. *New Technologies in Medieval and Renaissance Studies, Tempe 2012 (Medieval and Renaissance texts and studies 426; New technologies in medieval and Renaissance studies 3)*.
- Pazienza, M. T., Scarpato, N., Stellato, A., & Turbati, A.** (2012). Semantic turkey: a browserintegrated environment for knowledge acquisition and management. *Semant. Web J.* 3(3), 279–292.
- Piotrowski, M.** (2012). Natural Language Processing for Historical Textes. San Rafael, CA. (Synthesis Lectures on Human Language Technologies).
- Plachta, B.** (2018). Der ›edierter‹ Text: Grundpfeiler der Edition oder ›Zugeständnis‹ an den Leser? *Digitale Metamorphose: Digital Humanities und Editionswissenschaft*. Ed. by Kamzelak, R. S., & Steyrer, T. (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 2). text/html Format. DOI: [10.17175/sb002_002](https://doi.org/10.17175/sb002_002)
- Pollin, C., & Vogeler, G.** (2017). [Semantically Enriched Historical Data. Drawing on the Example of the Digital Edition of the 'Urfehdebucher der Stadt Basel'](#). *WHiSe 2017. Workshop on Humanities in the Semantic Web, Proceedings of the Second Workshop on Humanities in the Semantic Web (WHiSe II), co-located with 16th International Semantic Web Conference (ISWC 2017)*. Ed. by Adamou, A., Daga, E., & Isaksen, L (CEUR-Workshop proceedings 2014), p. 27-32
- Postles, D.** (ed.) (2011). Stubbington Account rolls from c.1247 and others. Retrieved from <http://www.historicalresources.myzen.co.uk/STUBB/Oprelim.html>

- Poupeau, G.** (2006). De l'index nominum à l'ontologie. Comment mettre en lumière les réseaux sociaux dans les corpus historiques numériques?. In : *Digital Humanities 2006*. The First ADHO International Conference: Conference Abstracts. (Paris: Université Paris-Sorbonne), pp. 161-164.
- Rauscher, P., & Serles, A.** (eds.) (2008-2018). Der Donauhandel. Quellen zur österreichischen Wirtschaftsgeschichte des 17. und 18. Jahrhunderts. Retrieved from <http://www.univie.ac.at/donauhandel/>
- Sahle, P.** (2013): *Digitale Editionsformen*, Norderstedt (Schriften des Instituts für Dokumentologie und Editorik 7-9).
- Sahle, P.** (2008-2017): *A catalog of Digital Scholarly Editions*. Retrieved from <http://www.digitale-edition.de/>
- Sanderson, R., Ciccarese, P., & Sompel, H. Van de** (eds.) (2013). Open Annotation Data Model, 8. Febr. 2013. Retrieved from <http://openannotation.org/spec/core/>
- TEI-Community** (2010). RDF and TEI. <http://tei-l.970651.n3.nabble.com/RDF-and-TEI-XML-tt2346163.html>
- TEI-Community** (2014). TEI and RDFa. <http://tei-l.970651.n3.nabble.com/TEI-and-RDFa-was-Re-SAWS-and-LOD-was-Re-Cross-references-among-segs-in-TEI-tt4025195.html>
- Kirchner, T., Nova, A., Blüm, C., Schreurs, A., & Wübbena, T.** (eds.) (2008-2012). Sandrart, J. von: *Teutsche Academie der Bau-, Bild- und Mahlerey-Künste, Nürnberg 1675/1679/1680*. Retrieved from <http://ta.sandrart.net/>
- Schaller, H. M.** (2000). Handschriftenverzeichnis zur Briefsammlung des Petrus de Vinea, unter Mitarbeit von Bernhard Vogel. Hannover. (MGH Hilfsmittel 18).
- Schulz, C., Putnam Miller, P., Marrs, A., & Allen, K.** (2002). Careers for Students of History, Section: Historians in Editing and Publishing. Retrieved from <https://www.historians.org/jobs-and-professional-development/career-resources/careers-for-students-of-history/historians-in-editing-and-publishing>
- Steinecke, H.** (1982). Brief-Regesten. Theorie und Praxis einer neuen Editionsform. *Zeitschrift für Deutsche Philologie* 101, p. 199-210.
- Steiner, E., & Stigler, J.** (2014-2017). *GAMS and Cirilo Client: Policies, documentation and tutorial*. <http://gams.uni-graz.at/o:gams.doku>
- Stevens, M. E., & Burg, S. B.** (eds.). *Editing Historical Documents: A Handbook of Practice* (Altamira Press, 1997).
- Thaller, M.** (1980). Automation on Parnassus Clio - a databank oriented system for historians. *Historical Social Research* 5, 3, pp. 40-65.
- Thaller, M.** (1988). Gibt es eine fachspezifische Datenverarbeitung in den historischen Wissenschaften? Quellenbanktechniken in der Geschichtswissenschaft. *Geschichtswissenschaft und elektronische Datenverarbeitung*. Ed. by Kaufhold, K. H., & Schneider, J. Wiesbaden (Beiträge zur Wirtschafts- und Sozialgeschichte 36), p. 45-83.
- Thaller, M.** (1992). The Historical Workstation Project. *Histoire et Informatique*. Ed. by Josef Smets, Ve congrès 'History & Computing' 4-7 septembre 1990 à Montpellier, Montpellier, p. 251-260.
- Thaller, M.** (1993). *Kleio. A Database System*, St. Katharinen: #Scriptura Mercaturae# (Halbgraue Reihe zur Historischen Fachinformatik : Serie B 11).
- Thaller, M., & Buzzetti, D.** (2012). Beyond Embedded Mark-up. *Digital Humanities 2012*. Retrieved from <http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/beyond-embedded-mark-up.1.html>
- Veluwenkamp, J. W., & Woude, S. van der** (2009 ff.). *Soundtoll Registers Online*. <http://www.soundtoll.nl>

- Vogeler, G.** (2015). Digitale Edition von Wirtschafts- und Rechnungsbüchern. *Wirtschafts- und Rechnungsbücher des Mittelalters und der Frühen Neuzeit*. Ed. by Gleba, G., & Petersen, N. Göttingen 2015, p. 307-328.
- Vogeler, G.** (2015). Warum werden mittelalterliche und frühneuzeitliche Rechnungsbücher eigentlich nicht digital ediert? *Grenzen und Möglichkeiten der Digital Humanities*. Ed. by Baum, C. & Stäcker, T. Wolfenbüttel (ZfdG - Sonderband 1). Retrieved from http://zfdg.de/sb001_007
- Vogeler, G.** (2016). The Content of Accounts and Registers in their Digital Edition. XML/TEI, Spreadsheets, and Semantic Web Technologies. *Konzeptionelle Überlegungen zur Edition von Rechnungen und Amtsbüchern des späten Mittelalters*. Ed. by Sarnowsky, J. Göttingen, p. 13-41.
- W3C** (2013). *W3C Data Activity Building the Web of Data*. <https://www.w3.org/2013/data/>
- Waitz, G.** (ed.) (1892). Ex rerum Danicarum scriptoribus saec. XII. et XIII. Ex historiis Islandicis. Ex rerum Polonicarum scriptoribus saec. XII. et XIII. Ex rerum Ungaricarum scriptoribus saec. XIII, Hannover; (MGH Scriptorum 29). Retrieved from http://www.mgh.de/dmgh/resolving/MGH_SS_29
- Zala, S.** et al. (ed.) (1978-2018). *Diplomatische Dokumente der Schweiz*. Bern/Zürich. <https://www.dodis.ch/>