Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing

Schriften des
Instituts für Dokumentologie und Editorik

herausgegeben von:

Band 15

Schriften des Instituts für Dokumentologie und Editorik — Band 15

Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing

edited by

Elena Spadini, Francesca Tomasi, Georg Vogeler

2021

BoD, Norderstedt

Bibliografische Information der Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über **http://dnb.d-nb.de/** abrufbar.

Digitale Parallelfassung der gedruckten Publikation zur Archivierung im Kölner Universitäts-Publikations-Server (KUPS). Stand 5. Dezember 2021.

Herstellung und Verlag: Books on Demand GmbH, Norderstedt
ISBN: 978-3-7543-4369-2
Einbandgestaltung: Stefan Dumont nach Vorarbeiten von Johanna Puhl und Katharina Weber
Satz: LuaTEX, Bernhard Assmann

Contents

Projects and Editions

Appendices

Infrastructures and Technologies

SPEEDy. A Practical Editor for Texts Annotated with Standoff Properties

Iian Neill, Desmond Schmidt

Abstract

Standoff properties can be used to record textual properties or annotations that may freely overlap and need not conform to a context-free grammar. In this way they avoid the ‘overlapping hierarchies’ problem inherent in markup languages like XML. Instead of embedding markup tags directly into the text stream, standoff properties are stored separately, and refer to positions in the text where each property starts and ends. However, this has the effect of tightly binding the properties to the text, and hence any change in the underlying text invalidates them. This limitation usually makes this method impractical in cases where the text is mutable, and is mostly used when the text is already fixed or proofread to a high standard. However, if it did become feasible to use standoff properties on mutable texts, this method could also be used in the process of text production, on dynamically evolving texts, such as emails, forum messages, personal notes and even drafts of academic papers. Digitised transcriptions of historical documents, whether produced manually or through OCR, could then be easily corrected at an earlier stage of typographic correctness. By overcoming the overlapping hierarchies problem this technique thus offers the prospect of significant productivity gains for producing digital editions, as well as a new mode of engagement for annotation. This paper describes the SPEEDy editor, a practical realisation of this technique. It outlines the editor’s foundational concepts, its standoff properties model, and its main interface features.

1 Introduction

SPEEDy (Neill 2020a) was created as part of a personal project to transcribe and annotate the letters of Michelangelo (1474–1564). As this project progressed, it became apparent that managing the numerous references to the people, places, and events mentioned in the letters required not only an annotation system, but also an editor, to allow for the updating of the content, its formatting, and annotations, in one place.

Annotations are inherently overlapping. Ranges of text may be arbitrarily selected by the user and annotations applied to them. However, the underlying HTML markup of a web page *is* strictly nested and forbids overlap. This creates a potential conflict

Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing. Ed. by Elena Spadini, Francesca Tomasi and Georg Vogeler. Schriften des Instituts für Dokumentologie und Editorik 15. Norderstedt: Books on Demand, 2021. 45–54.

between the two systems of markup. One is internal, describing layout and text structure, the other external or out-of-line, which points to it. Standard systems for specifying external annotations, such as the Open Annotation Model and annotation clients based on it, Recogito (n.d.) and Hypothesis.is (2011-), are limited in that they cannot specify arbitrary ranges within HTML (Sanderson et al., 2013, 2.1.4). They also assume that the text being annotated is immutable, since any change to it would potentially invalidate the annotations that refer to it (Tennison 2012, 7.2). Internal annotation systems such as RDFa (Herman et al. 2015) and Microformats (2020) are not subject to this limitation; however, they can only provide attributes for correctly nested HTML elements, and are hence unable to represent overlapping structures.

However, the distinction between the markup of the text being annotated and that of the annotations is artificial, since even formatting markup is partly semantic, and is thus a kind of annotation. For example, a paragraph element in HTML is not merely a formatting construct (e.g., left-aligned or with double line-spacing etc.) but also indicates that the enclosed text is a separate sequence of thoughts, distinct from the paragraphs that precede or follow it.

It has long been known that the transcription of historical sources often requires the description of overlapping structures (Renear et al., 1993). In markup languages like SGML and XML a context-free grammar is defined to regulate the syntax of the tags and text, the so-called DTD or schema. But this requirement, perfectly suited to the composition of new business documents, leads to problems when transcribing historical texts. For example, in holographs, revisions such as the joining of two paragraphs, overlapping underlining, and the conflicting demands of semantic, formatting and sentence structures, suggest that an approach seeking to redefine all markup as a kind of annotation might solve these conflicts in a new way. Buzzetti (2002) argued some time ago that *weakly embedded markup*, by which he seems to mean a kind of standoff annotation describing possibly overlapping ranges in the underlying text, may provide a better representation of textual structure and meaning than strongly embedded markup languages like XML. However, to be practicable, this would mean that the underlying text of the annotations would have to be editable and fluid, not fixed and immutable, and that the pointers into the text by the annotations, like the text itself, would have to be dynamically updateable.

SPEEDy differs from standoff markup editors like MUP (Glass & Di Eugenio 2002), the MATE annotation workbench (McKelvie et al., 2001), or the repository approach to standoff markup used in *Knora* (Rosenthaler et al., 2019), all of which preserve the hierarchical structure of the original embedded markup, and require an immutable base text. It also differs from web-based tools such as T-PEN (n.d.) for manuscript transcription, or editors like Quill (n.d.), which focus on editing, rather than annotation. SPEEDy instead combines features of both web-based editors and annotation clients in a single design.

The rest of this paper is divided into five sections. The first discusses the model for representing text in the editor, the second describes the data model for annotations, and the third describes the key features of the user interface. The fourth section describes future plans for the editor and the fifth draws some conclusions for its use in digital humanities projects.

2 Representing Text

SPEEDy is built on the notion of *cells,* which are individual characters arranged in a linked list, which forms the text stream. The concept of the text stream itself is expressed in two modes: *in-stream* and *out-of-stream* cells. An *in-stream cell* contains a single character and the sequence of *in-stream* cells collectively forms the plain text output. An *out-of-stream* cell is content which is visible in the editor panel, but which is not included in the plain text output. The assignment of text to one of these two modes is a matter of choice by the human editor. For example, a page number may or may not be considered part of the plain text output. Other examples of *out-of-stream* text include footnote labels, end-of-line hyphenation, and Leiden Convention symbols.

Since SPEEDy is a web-based editor, it uses the `<span>` element of HTML to implement individual cells, one character per cell. Any properties that the text acquires when loaded or during editing are stored in an array of standoff property JavaScript objects, which are linked to the `<span>` elements at the start and end of the annotation range. Therefore, annotations can move about naturally as the text is edited. Blocks of text such as paragraphs and other textual divisions are represented using the `<div>` element. No other HTML elements are used, or needed. Originally SPEEDy only supported `<spans>`s, linked together to form the text content. This kept the editor simple, and allowed the easy application of overlapping annotations. However, the need for paragraph level formatting such as left, right, centre, or justified text alignment, and indentation required the segmentation of this simple linked-list model of text into blocks. Navigation between blocks is now performed not by directly navigating the linked list structure, but via a tree-walking algorithm, which traverses the sequence of text cells as if it were contiguous. In other words, every index in the text refers to a unique location, either between two cells, or at the start or end of the entire text.

This distinction between blocks and characters is based on the assumption that, while semantic annotations are not inherently hierarchical, annotations describing layout require some kind of shallow hierarchy for representation in real-world formatting systems like HTML.

In the database the *in-stream* text is stored as plain text, with no markup of any kind. The *out-of-stream* text, and the markup which was contained in the attributes

```
{
  "text": "6th August. A thunderbolt fell on the cupola, but hardly did any damage.",
  "properties": [
    {
      "index": 0,
      "guid": "42d3333e-cfb7-4bd0-a344-3372b79a1d47",
      "type": "line",
      "text": "1",
      "startIndex": 0,
      "endIndex": 48,
      "attributes": {},
      "isDeleted": false,
      "userGuid": "fb067f75-a121-47c1-8767-99271c75cfc0"
    },
```


Figure 1. JSON export of SPEEDY annotation.

of the spans and blocks during editing, is stored as a set of properties, which point to the text in accordance with the standoff properties model. When a document is called up for editing again, the text and properties are recombined to generate the blocks and spans, so completing the cycle.

3 Representing Annotations

As described above, SPEEDy represents both formatting markup and semantic annotations as standoff properties that may freely overlap (Schmidt 2016). The more familiar term standoff markup refers to markup tags that represent a strict hierarchical text structure, conforming to a context-free grammar, which have been removed from the text and stored separately. Standoff properties, on the other hand, do not conform to a grammar, and do not have an intrinsic hierarchical structure. In XML, the data model for tags consists of elements and their attributes. Likewise, in SPEEDy, the standoff properties have their own data structure, which is described in this section.

One way to see the structure of annotations is to export a SPEEDy document to a JSON format. This shows how the document is represented as a plain text string, and its properties as an array. Figure 1 shows an example of this output, showing a short text and one of its standoff properties.

The essential fields of each property are:

- `type`. This represents the name of the annotation. In XML this would be the element name.
- `startIndex`. An integer representing the index position of the first character of the text range: $0 \leqq i < n$, where i is the index and n is the length of the text.
- `endIndex`. An integer representing the index position of the last character of the text range: `startIndex` $\leqq i < n$, where i is the index and n is the length.
- Not all annotations will describe a simple range of text. Two other possibilities are annotations that refer to a specific point in the text, and those that refer to the document as a whole.

- A null value for `endIndex` signifies that the property represents a point between two indexes, called here a *zero-width annotation* (ZWA).
- A null `startIndex` signifies that the property is metadata associated with the text as a whole, and does not refer to any specific position or range within it.

The other main fields of an annotation are:

- `value`. Any value which can be assigned to an annotation, depending on its function. For example, an entity annotation (i.e., for some kind of name) would use `value` to store the entity's identifier, while a font annotation would store the font name. There is no native *entity* or *font* annotation in SPEEDy, since `value` refers to something that is externally defined.
- `attributes`. A schema-less array field that is functionally equivalent to the attributes of an XML element. The choice between `value` and `attributes` for storing annotation data is one of convenience: typically, if it only makes sense for an annotation to have a single value then it is better to use `value`; if multiple values are required, then `attributes` should be preferred. An example of the latter is a Part of Speech annotation, which might contain data like gender, case, person, etc. Attributes are also used in LMNL, another markup formalism that uses standoff properties (Piez, 2015).
- `text`. For regular annotations of text ranges, the `text` field contains a copy of the text in its range in the text proper (arbitrarily truncated at 100 characters by default, but overridable). This facilitates reading the data either in the JSON output or in the data store. But the true object of the annotation is always the sequence of characters in the text starting at `startIndex` and ending at `endIndex`. For zero-width annotations, this field is used to store the *out of stream* text that denotes the inter-index position of the ZWA. For example, if a footnote were modelled as a ZWA, the footnote number or label would be stored in the `text` field.
- `guid`. The standoff property's unique identifier. No format is specified, but supplying an identifier for each annotation facilitates storage and querying in the external annotation store.

The `index`, `isDeleted` and `userGuid` fields are used internally by SPEEDy, and need not be explained here.

4 The Editor User Interface

The user interface is important because it is the window through which the user sees and manipulates the underlying textual and annotation data models. As explained above, text formatting and semantic annotations are merged in SPEEDy, and this

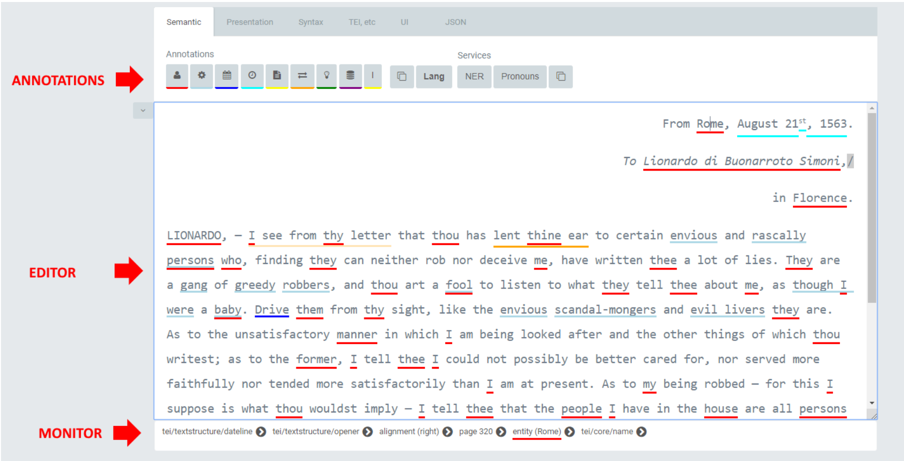


Figure 2. Screenshot of SPEEDy as integrated into Codex.

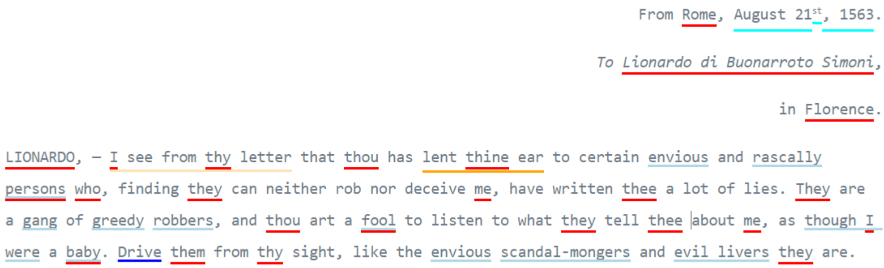
increase in cognitive load must be carefully handled so as not to overwhelm the user with too much detail (Nielsen 1993, 129f).

The annotations available in SPEEDy will vary greatly from one project to another, and are therefore fully configurable via a JSON file. This can include other vocabularies such as TEI (Neill 2020a). Although SPEEDy is an independent editor, it can be integrated in a variety of applications. The following screenshots are taken from the integration of SPEEDy in the Codex environment (Neill & Kuczera 2019), which provides visualisations of annotations across a document collection. It is based on a graph meta-model stored in Neo4j. However, all the visualisations provided below refer only to SPEEDy and do not concern the Codex system.

The editor interface is divided into three main sections: the annotation toolbar, the editor window and the monitor toolbar, as shown in Figure 2 (Carden 1913, 320; Neill 2020b).

Figure 3 shows a close-up of the annotated text of the above Michelangelo letter (Carden 1913, 320). Semantic annotations are visualised with coloured lines which are positioned below the line so that overlaps between properties are obvious. Also it is clear that the annotations can coexist perfectly with textual properties such as the right aligned text of the opening section of the letter.

Figure 4 shows the monitor toolbar, which changes whenever the user clicks on an annotation in the editor window. It displays any annotations that enclose the current cursor position. In Figure 2 the cursor is on the word Rome, which is a named entity that is itself inside several other annotation ranges. Hovering over an annotation in



From Rome, August 21st, 1563.

To Lionardo di Buonarroto Simoni,

in Florence.

LIONARDO, – I see from thy letter that thou has lent thine ear to certain envious and rascally persons who, finding they can neither rob nor deceive me, have written thee a lot of lies. They are a gang of greedy robbers, and thou art a fool to listen to what they tell thee about me, as though I were a baby. Drive them from thy sight, like the envious scandal-mongers and evil livers they are.

Figure 3. Detail of a text in SPEEDy showing the use of coloured underlines for visualising overlapping semantic annotations.

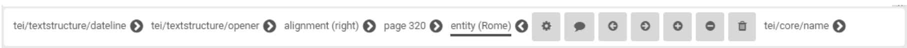


Figure 4. The monitor bar.

the monitor toolbar highlights both the annotation name and its range in the text. To the right of each annotation is a right-facing chevron which, when clicked, displays a list of buttons for editing the annotation (Neill 2020b). Figure 5 shows this in more detail.

Counting from left to right the buttons perform the following functions: [1] launches a window for changing the entity reference; [2] enables a comment to be added to the annotation (the comment itself is a standoff property document); [3] shifts the annotation to the left by one character; [4] shifts the annotation to the right by one character; [5] expands the annotation by one character; [6] retracts the annotation by one character; and [7] deletes the annotation. Although these interactions are fairly complex the buttons only appear when the right chevron is clicked, and also only apply to the currently selected annotation.

Figure 6 (Landucci & del Badia 1927, 299; Neill 2020b) shows two examples of out-of-stream text. Recalling the discussion in section 1 above, out-of-stream text is visible in the editor but is not included in the plain text output. The first example is a zero-width hyphen annotation which appears in the middle of the word *paternoster*.

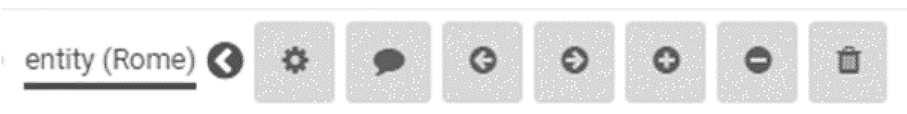


Figure 5. Controls in the monitor for manipulating an annotation.

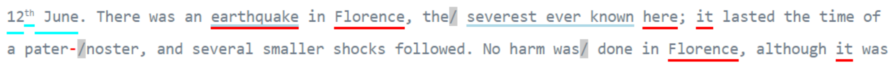

12th June. There was an earthquake in Florence, the/ severest ever known here; it lasted the time of
a pater-/noster, and several smaller shocks followed. No harm was/ done in Florence, although it was

Figure 6. Examples of *out-of-stream* text.

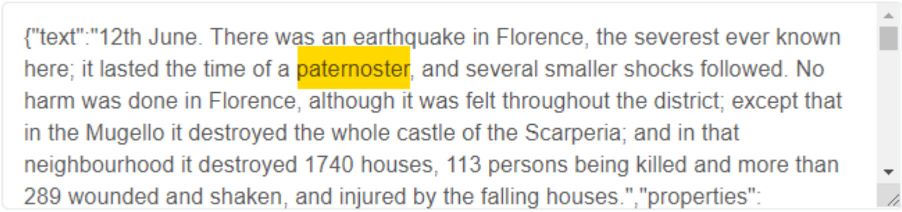

{"text":"12th June. There was an earthquake in Florence, the severest ever known here; it lasted the time of a paternoster, and several smaller shocks followed. No harm was done in Florence, although it was felt throughout the district; except that in the Mugello it destroyed the whole castle of the Scarperia; and in that neighbourhood it destroyed 1740 houses, 113 persons being killed and more than 289 wounded and shaken, and injured by the falling houses.","properties":

Figure 7. Example showing that the zero-width soft hyphen annotation in 'pater-noster' is not rendered in the output text stream.

The second is the line-break annotation, which is shown as a greyed-out forward slash. A hyphen annotation is used to represent a soft hyphen, which appears in the source but has only typographical, not linguistic, significance (which would be a hard hyphen). Including soft hyphens in the text content complicates the process of exact text matching, syntax analysis and lemmatisation. By modelling the hyphen annotation as a zero-width annotation, it can be displayed at its anchor location as an out-of-stream red-coloured hyphen.

Figure 7 (Landucci & del Badia 1927, 299; Neill 2020b) shows that the word *pater-noster* is not hyphenated in the recorded plain text section of the JSON output for this document. The same technique of using out-of-stream text is also applied to the line-break annotations.

5 Future Plans

Further developments planned for SPEEDy include:

- the ability to import and export TEI-XML text;
- support for scripts with contextual letterforms, such as Arabic and Syriac;
- support for Unicode planes other than the basic multilingual plane;
- the virtualisation of the editor window memory space to handle texts of any size;
- additional monitor sections for linking the editor to real-time visualisations;
- examination of the practicality of annotating both plain text and XML structures; and
- handling of multi-version and multi-layered texts, such as edited holographs.

6 Conclusion

SPEEDy has demonstrated the practicality of an editor based on standoff properties. Offering the user the possibility of editing texts that can be enriched with a customisable set of textual properties and annotations, and yet saved in a form that cleanly separates them from the base text, provides greater flexibility in the tasks of transcription and annotation. Corrections to OCR-scanned editions or transcribed original sources can now include markup and annotation from the start, rather than be introduced at a later stage in the production process.

Bibliography

Anon, 'Knora Documentation', *Knora* <https://docs.knora.org/>

Buzzetti, Dino, 'Digital Representation and the Text Model', *New Literary History*, 33 (2002), 61–88

Carden, Robert W., *Michelangelo Buonarroti: Michelangelo: A Record of His Life as Told in His Own Letters and Papers* (London: Constable & Co, 1913)

Glass, Michael, and Barbara Di Eugenio, 'MUP: The UIC Standoff Markup Tool', in *Proceedings of the Third SIGdial Workshop on Discourse and Dialogue*, SIGDIAL (Association for Computational Linguistics, 2002), 37–41 <https://doi.org/10.3115/1118121.1118126>

Herman, Ivan, Ben Adida, Manu Sporny, and Mark Birbeck, 'RDFa 1.1 Primer - Third Edition Rich Structured Data Markup for Web Documents', 2015 <https://www.w3.org/TR/xhtml-rdfa-primer/>

Hypothesis, founded by Dan Whaley, [2011-] <https://web.hypothes.is/>

Landucci, L, and I del Badia, *A Florentine Diary from 1450 to 1516. Trans. by Alice de Rosen Jervis* (New York: J.M. Dent, 1927)

McKelvie, David, Amy Isard, Andreas Mengel, and Morten Baum Møller, 'The MATE Workbench –an Annotation Tool for XML Coded Speech Corpora', *Speech Communication*, 33.1–2 (2001), 97–112

'Microformats', 2020 <http://microformats.org/>

Neill, Iian, 'Editor Demo', 2020a <https://argimenes.github.io/standoff-properties-editor/>

———, 'Standoff-Properties-Editor', 2020b <https://github.com/argimenes/standoff-properties-editor>

Neill, Iian, and Andreas Kuczera, 'The Codex – an Atlas of Relations', *ZfdG - Zeitschrift Für Digitale Geisteswissenschaften*, Sonderband 4 (2019) <http://www.zfdg.de/sb004_008>

Nielsen, Jakob, *Usability Engineering* (Boston: Academic Press, 1993)

Piez, Wendell, 'TEI in LMNL: Implications for Modeling', *Journal of the Text Encoding Initiative*, 8 (2015) <https://journals.openedition.org/jtei/1337>

'Quill', *Quill* <https://quilljs.com/>

Renear, Alan, Elli Mylonas, and David Durand, 'Refining Our Notion of What Text Really Is: The Problem of Overlapping Hierarchies', 1993 <http://cds.library.brown.edu/resources/stg/monographs/ohco.html>

Rosenthaler, L., B. Geer, and T. Schweizer, ‘Knowledge Organization, Representation, and Annotation’, *Knora*, 2019 <https://www.knora.org/>

Sanderson, Robert, Paolo Ciccarese, and Herbert Van de Sompel, ‘Open Annotation Data Model’, *Open Annotation Data Model*, 2013 <http://www.openannotation.org/spec/core/>

Schmidt, Desmond, ‘Using Standoff Properties for Marking-up Historical Documents in the Humanities’, *It – Information Technology*, 58.H.2 (2016), 63–59

Recogito, developed by Elton Barker, Leif Isaksen, Rebecca Kahn, Rainer Simon, and Valeria Vitale, n.d. <https://github.com/pelagios/recogito2>

Tennison, Jeni, ‘Best Practices for Fragment Identifiers and Media Type Definitions’, 2012 <https://www.w3.org/TR/fragid-best-practices/>

‘T-PEN Version 2.8’, *T-PEN* <http://www.t-pen.org/TPEN/>

Formal Models

Projects and Editions

Appendices

Biographical Notes

Thomas Ahrend (University of Basel, Switzerland – thomas.ahrend@unibas.ch) studied Musicology, Philosophy and Literary Studies in Frankfurt a. M. and Berlin. He received his MA 1996, and his PhD 2005 at Technische Universität Berlin with a dissertation on the instrumental music of Hanns Eisler. 1997–2010 member of the editorial staff of the Hanns Eisler Gesamtausgabe in Berlin. Since September 2010, member of the editorial staff of the Anton Webern Gesamtausgabe at Musikwissenschaftliches Seminar at University of Basel.

Peter Boot (Huygens ING, The Netherlands – peter.boot@huygens.knaw.nl) studied mathematics and Dutch language and literature; he wrote his PhD thesis about annotation in scholarly digital editions and its implications for humanities scholarship. He oversaw the creation of the digital edition of the letters of Vincent van Gogh. He is employed as a senior researcher at the Huygens Institute for the History of the Netherlands where he works, among other things, as a consultant in several edition projects.

Manuel Burghardt (University of Leipzig, Germany – burghardt@informatik.uni-leipzig.de) is head of the Computational Humanities Group at Leipzig University. He is interested in the use of digital tools and computational techniques to explore new modes of doing research in the humanities. His most recent areas of research are Sentiment Analysis in the Humanities, Drametrics, Computational Intertextuality, Computational Analysis of Movies and Series and Music Information Retrieval.

Toby Burrows (University of Oxford, United Kingdom – toby.burrows@oerc.ox.ac.uk) is a Senior Researcher in the Oxford e-Research Centre at the University of Oxford, and a Senior Honorary Research Fellow in the School of Humanities at the University of Western Australia.

Hugh Cayless (Duke University, USA - hugh.cayless@duke.edu) is Senior Digital Humanities Developer at the Duke Collaboratory for Classics Computing. Hugh has over a decade of software engineering expertise in both academic and industrial settings. He also holds a Ph.D. in Classics and a Master's in Information Science. He is one of the founders of the EpiDoc collaborative and currently serves on the Technical Council of the Text Encoding Initiative.

Hans Cools (University of Basel, Switzerland – 1961-2021) had a master degree in medicine and a specialization in orthopaedic surgery and traumatology (Universities of Ghent and Antwerp, Belgium, 1997), a bachelor's degree in physical therapy, and a standalone degree in informatics (1999). Through various research and project management positions, in both companies and academic institutions, he gained expertise in different aspects of the Semantic Web technologies, focusing particularly on formal data modeling and machine reasoning. Those positions were in internationally collaborative research projects in a biomedical setting, mainly of the 5-7th EU Framework Program. Foremost in these projects were semantic interoperability and reusability of data. Since 2016, he worked in the humanities, as knowledge engineer, ontologist, and Semantic Web technology expert, at the University of Basel, as part of the NIE-INE project, which highlights scholarly editing. He (co-)published several articles, and gave workshops on the implementation of Semantic Web technologies in biomedicine and the humanities. He passed away in April 2021.

Francesca Giovannetti (University of Bologna, Italy – francesc.giovannett6@unibo.it) is a second-year PhD student in Digital Humanities at the Department of Classical Philology and Italian Studies, University of Bologna. She received an MA in Digital Humanities from King's College London and a second cycle degree in Digital Humanities and Digital Knowledge from the University of Bologna. She is interested in combining digital scholarly editing with semantic web technologies and in the use of digital technologies in education.

Matthew Holford (University of Oxford, United Kingdom – matthew.holford@bodleian.ox.ac.uk) is Tolkien Curator of Medieval Manuscripts at the Bodleian Library, University of Oxford.

Marijn Koolen (Royal Netherlands Academy of Arts and Sciences - Humanities Cluster, The Netherlands – marijn.koolen@gmail.coml) studied artificial intelligence and wrote his PhD thesis on using hyperlinks in information retrieval algorithms. He has worked on scholarly annotation for digital humanities research and on annotation-related information behaviour and information systems. He works as a researcher and developer at the Humanities Cluster of the Royal Netherlands Academy of Arts and Sciences, where he leads a project on developing annotation support within the *CLARIAH research infrastructure* project.

David Lewis (University of Oxford, United Kingdom – david.lewis@oerc.ox.ac.uk) is a Research Associate in the Oxford e-Research Centre at the University of Oxford.

Andrew Morrison (University of Oxford, United Kingdom – andrew.morrison@bodleian.ox.ac.uk) is a Software Engineer in the Bodleian Digital Library Systems and Services, Bodleian Library, University of Oxford.

Stefan Münnich (University of Basel, Switzerland – stefan.muennich@unibas.ch) studied musicology and communication science at the Technische Universität Berlin, MA 2011 with a thesis on cantional setting in Heinrich Schütz's Becker-Psalter. 2012 research assistant, 2013–2015 research associate of the Felix Mendelssohn Bartholdy. Sämtliche Briefe edition at University of Leipzig (co-editor of vols. 9 & 12). Since October 2015 research associate of the Anton Webern Gesamtausgabe, Basel; received his Doctorate degree in 2020 at the department of musicology at the University of Basel with a dissertation about music notation and its codes.

Iian Neill (Digital Academy of the Academy of Sciences and Literature, University of Mainz - Iian.Neill@adwmainz.de) is a visiting researcher at the Digital Academy of the Academy of Sciences and Literature Department at the University of Mainz, Germany. He is the creator of Codex, a text annotation environment which uses standoff property annotation to generate entities in a graph meta-model. Codex is currently being used to produce a digital edition of the epistles of Hildegard von Bingen at the Digital Academy in Mainz.

Roberta Padlina (University of Basel, Switzerland – roberta.padlina@unibas.ch) studied medieval philosophy at the University of Fribourg, Switzerland, obtaining a doctoral degree in June 2020. She has twelve years of professional experience in the field of Digital Humanities, thanks to which she has been able to work closely with different actors involved in the online publication of open access research. Roberta has worked for several years for e-codices –Virtual Library of Manuscripts in Switzerland and currently coordinates the National Infrastructure for Editions (NIE-INE) project. Roberta's main focus is on the opportunities and challenges that the digital shift poses for traditional education and research institutions, including developing semantic web strategies for scholarly publications and cultural goods.

Kevin Page (University of Oxford, United Kingdom – kevin.page@oerc.ox.ac.uk) is a Senior Researcher in the Oxford e-Research Centre and Associate Member of Faculty in the Department of Engineering in the University of Oxford.

Miller C. Prosser (University of Chicago, USA – m-prosser@uchicago.edu) earned his Ph.D. in Northwest Semitic Philology from the University of Chicago. His academic interests include the social and economic structure of Late Bronze Age Ras Shamra-Ugarit and the use of computational methods for philological and archaeological research. Miller is the Associate Director of the Digital Studies MA program at the University of Chicago where he teaches courses on Data Management and Data Publication for the Humanities. He also works as a researcher at the OCHRE Data Service of the Oriental Institute of the University of Chicago where he consults with and supports research projects using the Online Cultural and Historical Research Environment (OCHRE). He has also worked as a tablet photographer for the Mission de Ras Shamra (Ugarit) and the Persepolis Fortification Archive Project, employing advanced digital photographic methods such as reflectance transformation imaging, photogrammetry, and high-resolution digital scanning.

Matteo Romanello (Université de Lausanne, Switzerland - matteo.romanello@unil.ch) is Ambizione SNF Lecturer at the University of Lausanne, where he conducts a project on the commentary tradition of Sophocles' Ajax. Matteo is a Classicist and a Digital Humanities specialist with expertise in various areas of the Humanities, including archaeology and history. After obtaining his PhD from King's College London, he worked as a research scientist at EPFL's DHLAB on the Linked Books and Impresso projects, before moving to his current position. He was also teaching fellow at the University of Rostock, researcher at the German Archaeological Institute, and visiting research scholar at Tufts University.

Sandra Schloen (University of Chicago, USA – sschloen@uchicago.edu) is the Manager of the OCHRE Data Service at the Oriental Institute of the University of Chicago, and is the co-designer and developer of the Online Cultural and Historical Research Environment (OCHRE). Trained in computer science and mathematics (B.Sc. University of Toronto; M.Ed. Harvard University), Sandra has spent over 30 years working with technology as a systems analyst, technical trainer, and software developer. A long association with colleagues in the academic community has enabled her to develop a specialty in solving problems in the Digital Humanities where challenges of data capture, data representation and data management abound. Specifically, she has served extensively as a database manager for several archaeological projects in Israel and Turkey, and supports a wide range of research projects at the Oriental Institute and at other universities.

Desmond Schmidt (University of Bologna - desmond.allan.schmidt@gmail.com) has a background in classical Greek philology, information security and eResearch. He has worked on several scholarly edition projects, including the Vienna Wittgenstein Edition (1990–2001), Digital Variants (2004–2008), the Australian Electronic Scholarly Editions project (2012–2013), the Charles Harpur Critical Archive (2014-) and a pilot edition of Gianfrano Leopardi's Idilli (2018-). He currently works on developing practical web-based tools for making, visualising and publishing digital scholarly editions.

Colin Sippl (University of Regensburg, Germany – colin.sippl@ur.de) is currently a project employee at the University Library of Regensburg. Since 2017, he has been working on extending the open access services of the Electronic Journals Library (EZB). More recently, he has started developing and setting up a digital repository for literature, artefacts and experiments relating to the early life sciences based on the Invenio framework. He specialised in textual data mining and the development of media services in the institutional domain.

Elena Spadini (University of Lausanne - elena.spadini@unil.ch) is a postdoctoral researcher at the University of Lausanne. She holds a Ph.D. in Romance Philology from the University of Rome Sapienza (2016) and a M.A. in Digital Humanities from the École nationale des chartes (2014). She was a Marie Curie fellow in the IT Network DiXiT and co-directed the related volume Advances in Digital Scholarly Editing (Sidestone Press, 2017). She published in international journals and taught specialized courses in various European countries in the field of Digital Philology.

Francesca Tomasi (University of Bologna - francesca.tomasi@unibo.it) is associate professor in Archival Science, Bibliography and Librarianship at the University of Bologna (Italy). Her research is mostly devoted to digital cultural heritage, with a special attention to documentary digital edition, and a focus on knowledge organization methods in archives and libraries. She is member of different scientific committees of both associations and journals. In particular, she is President of the Library of the School of Humanities in the University of Bologna (BDU - Biblioteca di Discipline Umanistiche), Director of the international second cycle degree in Digital Humanities and Digital Knowledge (DHDK), President of the Italian Association of Digital Humanities (AIUCD – Associazione per l'Informatica Umanistica e la Cultura Digitale), and co-head of the Digital Humanities Advanced Research Center (/DH.ARC). She wrote about 100 papers and 4 monographs related to DH topics. She is editor and scientific director of several digital scholarly environments.

Athanasios Velios (University of the Arts London, United Kingdom – a.velios@arts.ac.uk) is Reader in Documentation at the University of the Arts London.

Georg Vogeler (University of Graz - georg.vogeler@uni-graz.at) is professor for Digital Humanities at the University of Graz and scientific director of the Austrian Center for Digital Humanities and Cultural Heritage at the Austrian Academy of Sciences. He is a trained historian (Historical Auxiliary Sciences). He spent several years in Italy (Lecce, Venice). In 2011, he became member of faculty at the Centre for Information Modelling at Graz University, where he was nominated full professor for Digital Humanities in 2016 and head of department in 2019. His research interests lie in late medieval and early modern administrative records, diplomatics (digital and non digital), digital scholarly editing and the history of Frederic II of Hohenstaufen (1194–1250). He was and is part in several national and international research projects related to his research interests.

Christian Wolff (University of Regensburg, Germany – christian.wolff@ur.de) has been Professor of Media Informatics at the Institute for Information and Media, Language and Culture at the University of Regensburg since 2003. He holds a PhD in information science and is a habilitated computer scientist. His research interests include: human-computer interaction, multimedia and web-based information systems, (multimedia) software engineering and information retrieval (in particular information literacy and social media).