

The logo for 'ENCODING CULTURES' features the words in a bold, black, sans-serif font. The text is enclosed within a yellow L-shaped frame. Above and below the text are horizontal lines: three thin lines above and three thicker lines below. The year '2023' is highlighted in yellow in the text below.

**ENCODING  
CULTURES**

JOINT MEC AND TEI  
CONFERENCE **2023**

## **Book of Abstracts**

**4–8 September 2023**  
**Paderborn University, Germany**

Edited by Raffaele Vigiante, Johannes Kepper, Peter Stadler, Joachim Veit



# Table of Contents

From the Chair of the Program Committee	8
From the Local Organizers	10
Program Committee	12
Local Organizers	13

## Keynotes

Opening Keynote: Anna Kijas, "Preserving Cultural Heritage through Digital Activism and Community"	14
Closing Keynote: Till Grallert, "Mind the <gap/>s! Digital editing in a world in crisis"	15

## Abstracts

< ♫ > for Non-Western Scripts	17
3D Text Encoding and TEI: Text, Editions, and Spatiality	19
A (cautionary) tale of two texts	24
A digital annotated parallel corpus of Athonite text types: Gospel of Matthew	26
A multi-media dictionary of endangered languages with TEI Lex-0: a case study of Hatoma, Yaeyama Ryukyuan	31
A Preliminary Proposal for Digital Scholarly Editing that Uses Modern Japanese autograph manuscripts: How to markup autograph manuscripts of Rampo Edogawa	33
"Ain schone kunstliche Underweisung": Modelling German lute tablature in MEI	36
Alaine Locke's <i>New Negro Anthology</i> in TEI and MEI: A DEFCON Teaching Collaboration across Two Universities	38
An AI-assisted Digital Scholarly Editing System for Buddhist Studies.	40
An annotation model for software mentions and citations	43
An edition of the writing <i>Die Wagnerfrage</i> in the online portal of the Joachim Raff Archive	46
An MEI-TEI framework for the description of musical prints	48
Bellini Digital Correspondence meets MEI	50
Between bespoke customization and expressive interface: a reflection from the Frankenstein Variorum project	55

Big (Work) Data: Making and Using MEI for Composers' Digital Work Catalogues	58
Brainstorming on the Encoding of Eastern Neumes: Middle-Byzantine Notation as a Case-Study	60
Building sustainable infrastructure for scholarly communities	63
Building Workflows for HTR to TEI Up-Conversion and Enhancement	67
Comparison of MusicXML export capabilities of different scorers	69
Computing and displaying similarity in Beethoven's vocal sketches and score for <i>Leonore</i> (1805)	70
Cracking the Code: Overcoming the Challenges of Encoding Correspondence	72
Creating Digital Editions with EditionCrafter	74
Digital critical edition of Čiurlionis' piano music	76
Digital Representation of 'A Match of Crickets in Ten Rounds of Verse and Image': Text Encoding and Viewer Implementation for Japanese Poetry Match	79
Digitization of Ryukyuan Music: MEI of Kunkunshi Notation for Classical and Modern Songs with Sanshin	81
Digitizing Buddhist Genealogy: Encoding the Shinran Shonin Montei Kyōmyo-chō	83
Do <del>not</del> touch the <object>. Extending the <object> element based on the example of the Visual Archive Southeastern Europe	87
Encoding in human centered machine learning workflows: case study on mensural ligature recognition	90
Encoding mathematics: A concept for a data model for representing early modern and modern mathematical notations in Digital Editions	93
Encoding Notre Dame polyphony for corpus analysis	97
Encoding Orchestral Parts that Share a Staff in MEI: Guidelines and a Parts Extraction Tool	99
Encoding previously Transcribed Oral Literature: The Forager Folklore Database (FFDB)	102
Encoding Traditional Court Music and Dance Scores Using the Music Encoding Initiative	104
Encoding Transcultural Texts - Applying TEI to Early Modern Chinese Christian Literature	107

Exploring the Genesis of Complex Music Manuscripts	111
FAIR Derived Data in TEI for Copyrighted Texts	113
Five Aspects of ViFE	117
From TEI to CIDOC-CRM with LINCS XTriples: Preserving Meaning across formats?	120
Half-day tutorial: Music Encoding with mei-friend	122
“Herzlichst in Eile Ihr Mahler” – Indexing the letters of Gustav Mahler	125
Hybrid music editions – challenges at the Reger-Werkausgabe in retrospective.	126
Integrating TEI and MEI	129
Interactive music encoding with Verovio	131
Interoperability of Text and Image in the Digital Edition of the Wenceslas Bible	133
Introduction to publishing XML with static site and front-end technologies	135
Jazz music encoding for education: the collaboration between MEI and TEI.	138
Lessons from the Classroom: MEI for Data Scientists	141
Lights, Camera, Slugline? Encoding Screenplays in TEI	144
Managing Different (Encoding) Cultures when Building a TEI Corpus of Historical German Newspapers	146
Materiality in editions of 20th-century paperbound correspondence	150
MEI Basics	153
MEI encoding and analysis of “airs connus” across multiple genres	154
MEI Transformation for the Winterreise Dataset	156
mei-friend v1.0: Music Encoding in the Browser	159
Meta-Models Matter – The tscore R&D Project and “Die Kunst der Fuge”	161
Metadata in the DFG Viewer for Musical Sources: Requirements for the exchange of meta and structural data in MEI format	165
Mixing Modalities: Graphical and Text-Based Interaction in Music Notation Editing	166
Multicultural & Multilingual TEI Encoding: A Comparative Study of British Romanticism and Chinese Tang Poetry through the Lens of Romantic Ecology and Affect Narratology	170

Music-Text Interlinking as a Challenge for Digital Encodings of Music-Theoretical Writings	172
Networking Editions: TEI and the Semantic Web with TAG and RDF	174
Notation editor for transcribing Hindustani classical music Bandish	177
ODD ODDities: some problems with the ODD language unearthed by an attempt to write a processor	185
On the Relationship Between Interfaces and TEI in Digital Scholarly Editions	188
PDB18 - The German Letter in the 18th Century	191
Pedagogical Approaches to Encoding	195
Publishing TEI and MEI with TEI Publisher	198
Reconciling corpora from enthusiast and scholarly communities: a sustainable approach to augmenting the Electronic Corpus of Lute Music	200
Relation <sup>3</sup> : How to relate text describing relationships with structured encoding of the relationships?	203
Selective Encoding: Reducing the Burden of Transcription for Digital Musicologists	207
Studying Poetry through Music: The Tasso in Music Project	210
Teaching Encoding In and Out of the Classroom	212
TEI Lex-0: Recent Developments and New Directions	216
TEI Processing and Computational Analysis on Encoding Judgment Text	221
TEI semantic in graph-based digital scholarly editions	224
TEI XPointer Schemes - Implementation and Example Application	229
Text Encoding without //text: The use of //abstract as means to avoid the one-dimensionality of ego-networks in the 'Buber-Correspondences Digital' project	231
Text-genetic studies on Carl Maria von Weber's opera fragment "Die drei Pintos"	233
The Culture of Devotion – And what TEI Encoding tells us about it	237
The Past, Present, and Future of the Plaine and Easie Code	244
The Ricercar Data Lab score viewer	246
The WizKit: A New Environment for Editing and Presenting MEI Metadata	248

Thinking about collation	250
“This type of data in the header” – Encoding, (re)use and standardization of metadata in the context of TEI and MEI	254
Tools, Formats, Mappings: Encoding Franz Liszt’s Œuvre between MARC and MEI	257
Toward a TEI/RDF Encoding for Semantic Annotations: Concept and Implementation as LOD Editor	260
Towards correspSearch v3.0	264
Towards shared TEI model/s for institutional minutes and protocols – protokolleditionen.eu	268
Transitional MerMEId – Responding to Community Needs in Software and Sustainability	270
Travelling Humboldt–Data on the Move	272
Updates and Outcomes from the Guatemalan Digitization and Encoding Project	277
Using TEI for Chinese Architectural Data	280
Valid, but readable: writing D2d for constructing XML	282
Visual or Symbolic? Best Practices for Encoding Neumes	288





# From the Chair of the Program Committee

The “Encoding Cultures” conference brought together, for the first time, the Music Encoding Initiative (MEI) and Text Encoding Initiative (TEI) communities, both of which are involved in the digitization and encoding of cultural heritage artifacts. This conference served both as the 23<sup>rd</sup> annual meeting of the TEI community and the 11<sup>th</sup> annual Music Encoding Conference. While there are fundamental differences between these communities’ objects of study—musical versus textual artifacts—on a closer look one can find overlapping approaches in regards to data modeling, encoding theory, and digital publication that reveal the kinship of these two distinct *cultures of encoding*.

This kind of homecoming was an opportunity for taking a broader perspective on encoding practice. The Call for Proposals invited reflections on the complexities of *encoding different cultures*, encouraging contributions that engaged with multilingual and multicultural aspects of text and music encoding, such as multilingual metadata, support for non-Latin scripts and non-Western music notation, support for right-to-left languages, and interface internationalization. In other words, “Encoding Cultures” provided a space to acknowledge—within the encoding communities of practice—the important role that digital humanities scholars play in the “construction of the digital cultural record”, as emphasized by Roopika Risam in her book *New Digital Worlds (2019)*. In her postcolonial critique of the digital humanities, Risam highlights how “cultural canons are being reproduced and amplified not only in the visibility and discoverability of knowledge, but in epistemologies of digital knowledge production as well”. The awareness of this risk must inform our work as we as digital humanities scholars apply encoding as an epistemic and rhetorical practice, a point also echoed by the conference keynote speakers.

Till Grallert, in his keynote “[Mind the <gap/>s! Digital editing in a world in crisis](#),” explained how digital editing is entangled in cultural practices of what constitutes cultural heritage and invited us to consider “who is encoding cultures?” and “whose encoding cultures?”. Pinned against the backdrop of his work on early Arabic periodicals, Grallert offered “a practical critique of the epistemic violence exercised by digitisation efforts”. Anna Kijas, in her keynote “[Preserving Cultural Heritage](#)

[through Digital Activism and Community](#),” showed us the impact that digital preservation can have on today’s lives by addressing urgent needs and shaping current events. Kijas co-leads Saving Ukrainian Cultural Heritage Online (SUCHO). “Russia’s invasion of Ukraine on February 24th was a wake-up call for cultural heritage professionals worldwide. [...] (SUCHO) is a global grassroots initiative of more than 1,500 volunteers which formed as a response to the invasion to support the digital preservation of Ukrainian cultural heritage”. Her keynote showed how digital activism and community drive digital preservation, but also revealed the fragility of our collective digital cultural record and how important it is to plan for a more resilient future for digital cultural heritage protection.

The conference contributions covered a broad scope of disciplines, methods, and techniques across text and music encoding. As the Program Committee had hoped, there were several contributions that dealt specifically with overlaps between the two practices, from shared technological stacks to primary sources and corpora combining music and text. Contributions also reflected the communities’ multicultural focus and composition with papers on non-Western music notation such as Japanese ryukyuan music, encoding of right-to-left script systems, notation systems for Indian classical music, and more. This year’s review process was fully open, giving the opportunity to authors and reviewers alike to participate honestly and critically in the process. I found this to be particularly beneficial as two distinct but related communities came together. The 35-member Program Committee was evenly composed of music and text encoding experts and I thank them for all their work in shaping the conference and diligently reviewing proposals. In total, the Committee accepted 20 short papers, 35 long papers, 6 panels, 23 posters, and 8 pre-conference workshops.

The local organizers made the event comfortable and accessible to all participants and deserve the highest praise, particularly Johannes Kepper and Peter Stadler, who were supported by a large and truly dedicated team. Their commitment to a fully hybrid conference was a remarkable success, by bringing in remote attendees as much as possible, including through a dedicated projection of the virtual audience for all of the main conference sessions. I thank the conference’s multiple sponsors, particularly for making available student travel bursaries. Last but not least, thanks are due to all the contributors, who took to heart the challenges and opportunities posed by this interdisciplinary event and its Call for Proposals.

Raff Viglianti, Program Committee Chair

# From the Local Organizers

Fifteen years ago, we attended the TEI Conference at King's College London. At that conference, the TEI's Special Interest Group on Music met and discussed whether that group should become the home for music encoding. Because the community for music encoding had significant intersections with the TEI community, but also other relevant parts not represented by the TEI, eventually a separate Music Encoding Initiative evolved in the years to follow. Many relations were maintained between both communities, mostly on the level of individuals and shared technologies. It was a great pleasure for us that both communities accepted our invitation to meet for their first joint conference in Paderborn. Being actively involved in both TEI and MEI, we were quite confident that combining both events would be a stellar opportunity for interchange and mutual inspiration. However, it was absolutely clear that this required a great deal of flexibility on the part of all parties involved, beginning already with the time of year when the conference could take place. There were many more concerns like this, but the TEI Board and Council, the MEI Board and everyone else involved were very considerate, trying to find good compromises that were widely agreeable. We thank them for their trust and giving us the opportunity to host this conference.

With his experience in both communities, Raff Viglianti proved to be the ideal Program Committee chair that we expected him to be. Together with a very engaged program committee made up of members from both communities, he helped to shape this conference as a truly joint one, where music and text sessions would not sit side by side, but would instead provide a space where overarching questions would help bring together different perspectives, giving a much better overview of our different encoding cultures and highlight commonalities and diverging practices. We are very grateful for all the work that Raff and the Program Committee has invested.

We would also like to thank all the contributors who really brought this concept to life and gave the conference the thematic breadth we had hoped for. The book of abstracts at hand clearly shows how boundaries between scholarly disciplines are sometimes rather artificial, and that many seemingly marginal topics actually build bridges to other fields and thus represent a great enrichment for research. The same is certainly true for culture in general, which encourages us to take different perspectives on our topics. In that sense, the opening and ending keynotes from

Anna Kijas and Till Grallert proved to be an ideal combination, showing us the applicability of the data modelling expertise from digital humanists in real-world scenarios like the Russian invasion of the Ukraine, and how privileged our use of computing resources and biased our methods and workflows are, giving us all inspiration for how to improve our future work.

We would like to thank our sponsors, the German Research Foundation, the Stiftung Westfalen, and Bielefeld University Press. Finally, we would like to thank Paderborn University and its Musicological Seminar for their continued support. But, without the wonderful team of volunteers from the "Virtueller Forschungsverbund Edirom", it wouldn't have been possible to host this joint conference – something that we couldn't have dreamed of fifteen years ago.

Johannes Kepper and Peter Stadler

# Program Committee

- Clifford Anderson, Vanderbilt University, USA
- Helena Bermúdez Sabel, University of Neuchâtel, Switzerland
- Elisa Beshero-Bondar, Penn State Behrend, USA
- Benjamin W. Bohl, Goethe-University Frankfurt and Paderborn University, Germany
- Joy H. Calico, Vanderbilt University, USA
- Iacopo Cividini, Internationale Stiftung Mozarteum, Austria
- Constance Crompton, University of Ottawa, Canada
- James Cummings, Newcastle University, UK
- Sophia Dörner, SLUB Dresden, Germany
- Timothy Duguid, University of Glasgow, UK
- Jessica Grimmer, University of Maryland, USA
- Martin Holmes, University of Victoria, Canada
- Diane Jakacki, Bucknell University, USA
- Johannes Kepper, Paderborn University, Germany
- Anna E. Kijas, Tufts University, USA
- David Lewis, University of Oxford & Goldsmiths, University of London, UK
- Tiziana Mancinelli, Ca' Foscari University of Venice, Italy
- Wolfgang Meier, eXist Solutions, Germany
- Stefan Münnich, University of Basel, Switzerland
- Patricia O'Connor, Independent Scholar
- Frauke Pirk, University of Mainz, Germany
- Laurent Pugin, RISM Digital Center, Switzerland
- Kristina Richts, Paderborn University, Germany
- Gimena del Rio Riande, CONICET, Argentina
- Daniel Russo-Batterham, The University of Melbourne, Australia
- Sabine Seifert, University of Potsdam & Berlin-Brandenburg Academy of Sciences and Humanities, Germany
- Peter Stadler, Paderborn University, Germany
- Margrethe Støkken Bue, National Library of Norway
- Raffaele Vigiante, University of Maryland, USA (chair)
- David M. Weigl, mdw – University of Music and Performing Arts Vienna, Austria
- Laura Weakly, University of Nebraska-Lincoln, USA

# Local Organizers

- Martin Albrecht-Hohmaier
- Hizkiel Mitiku Alemayehu
- Nikolaos Beer
- Katharina Bergmann
- Benjamin W. Bohl
- Irmlind Capelle
- Anne Ferger
- Dennis Friedl
- Vera Grund
- Kristin Herold
- Dunja Jeske
- Daniel Jettka
- Johannes Kepper (chair)
- Elena Minetti
- Ran Mo
- Andreas Münzmay
- Salome Obert
- Kristina Richts-Matthaei
- Dennis Ried
- Daniel Röwenstrunk
- Mark Saccomano
- Agnes Seipelt
- Peter Stadler
- Joachim Veit
- Jan-Peter Voigt
- Anastasia Wawilow

# Opening Keynote: Anna Kijas, “Preserving Cultural Heritage through Digital Activism and Community”

Anna Kijas, Tufts University

## Opening Keynote

Russia’s invasion of Ukraine on February 24<sup>th</sup> was a wake-up call for cultural heritage professionals worldwide. For several decades, cultural institutions have engaged in digitization of cultural artifacts, in light of the invasion, however, it became clear that the servers on which these digital objects are stored are as vulnerable as the buildings which house the analog originals. Saving Ukrainian Cultural Heritage Online (SUCHO) is a global grassroots initiative of more than 1,500 volunteers which formed as a response to the invasion to support the digital preservation of Ukrainian cultural heritage. SUCHO began as an emergency web-archiving effort, preserving more than 50TB of cultural data from over 5,000 websites of Ukrainian museums, libraries, and archives. Since August of 2022, SUCHO, has expanded its efforts to donating digitization equipment to Ukrainian institutions, providing training, and supporting storage of and access to digitized and born-digital heritage.

In this keynote, Anna Kijas, co-founder of SUCHO, will share the story of SUCHO and explore how this initiative drew together a community of digital activists with a shared vision of ensuring that digitized and born-digital cultural heritage is pre-emptively protected in the future. She will discuss how SUCHO has promoted an ethos of community, collaboration, and open knowledge/resources amongst the volunteers, and in the relationships created with Ukrainian institutions and cultural heritage professionals. Much as World War II laid the groundwork for robust cultural heritage protection organizations that still operate today, the global cultural heritage community can learn from the invasion of Ukraine to plan for a more resilient future for digital cultural heritage protection.

# Closing Keynote: Till Grallert, “Mind the <gap/>s! Digital editing in a world in crisis”

Till Grallert, Humboldt-Universität zu Berlin

## Closing Keynote

Digital editing is a well-established craft with a decades-long history, widely-agreed upon sets of techniques and tools, with technical infrastructures, governing bodies, and striving communities of practice such as the TEI and MEI. At the core of the computational turn in the humanities, they provide access to the cultural record at an unprecedented scale. Yet, digital *editing* is entangled in cultural practices and much older politics of archives and preservation, of what constitutes cultural heritage and, ultimately, the politics of representation: "Who is encoding cultures?" and "whose encoding cultures?". *Digital* editing and its affordances depend on socio-technical infrastructures that are deeply rooted in capitalism and histories of exploitation of both human and natural resources. Such infrastructures of networked computational mediation range from standards of character encoding, to protocols of data exchange, legal regimes, and corporate ownership over vast swaths of the public sphere, thus reinforcing planetary power relations and reifying the cultural hegemony of the Global North and, more specifically, its American and anglophone constituents. Access to the digital means of production or rather the means of digital production is severely restricted for the societies and communities of the Global South. Even more, these societies and communities bear a disproportionate share of the high environmental and social costs and are often the first to suffer catastrophic consequences.

The keynote will explore these multi-layered digital divides embodied in the infrastructural underpinnings of modern scholarship against the backdrop of my work on early Arabic periodicals between the mid-nineteenth century and the end of World War I. Instead of a theoretical essay, I will offer a practical critique of the epistemic violence exercised by digitisation efforts through an account of two research projects, [Jarā'id](#) and [Open Arabic Periodical Editions](#) rooted in *minimal computing* and *pirate care* and focus on the always concrete affordances of the digital. Arabic—one of only six official languages of the United Nations with more than 420 million active speakers and the liturgic language of almost two billion Muslims as well as the world's second most important script after Latin—is a



(scandalous) prime example for under-resourced languages and systems of knowledge production. The textual record of Arabic periodicals from South-West Asia and North Africa as well as the global diasporas thus forces us to acknowledge and address the (not just linguistic) imperialism of global networked knowledge production.

# <Ⲁ> for Non-Western Scripts

Christian Lück, University of Münster, Germany

Short Paper

**Keywords:** TEI, non-western scripts, internationalization, ODD

## Abstract

Encoding right-to-left script in TEI-XML is a hassle unless one uses an editor that hides away the tags. The problem arises when element names in Latin script interrupt the right-to-left rendering on the editor screen by the Unicode Bidirectional Algorithm. (Davis et al. 2022) However, hiding the tags is not the only solution. With `<altIdent>` TEI offers a means for declaring alternative names for elements and attributes: names in another language and even in another script. Thus, we can have valid right-to-left-only TEI documents in XML version 1.1, which are readable and editable while tags are visible.

However, using `<altIdent>` quickly feels like introducing entropy to the schema. Therefore, the paper suggests not to translate, but to transliterate identifiers. Compared to translation, transliteration is a mechanical process. The coding effort is low. An implementation is at hand in the Unicode ICU library,<sup>[1]</sup> and there is even an XPath binding for this library,<sup>[2]</sup> so that transliteration is available in XSLT and XQuery. Thus, generating a full ODD with element and attribute aliases in Arabic, Syriac, Hebrew, etc. script becomes simple and fast.<sup>[3]</sup> It can be pre-built so that the hurdle for encoding non-western script in TEI becomes much lower. The `<Ⲁ>` from the paper's title is a `<TEI>` tag transliterated to Syriac script: `<altIdent xml:lang='en_Syrc'>Ⲁ</altIdent>`.

Transforming a TEI document with alternative names to its equivalent with names in Latin script is always possible through the mapping of identifiers and alternative identifiers deposited in the ODD. Direct re-transliteration of alternative names is only possible under certain preconditions.

The paper will first analyze the problem when bidirectional text is displayed on a screen. It will show how well-formed tags look in a right-to-left script. It will then introduce the transliteration of XML names in documents and ODDs. It will discuss challenges regarding attributes in the XML namespace, especially `@xml:id`.

## Bibliography

Davis, Mark et al. (2022): Unicode Standard Annex #9: Unicode Bidirectional Algorithm Latest version: <https://www.unicode.org/reports/tr9/>; Version 15.0.0: <https://www.unicode.org/reports/tr9/tr9-46.html>

## About the author

**Christian Lück** has a PhD in German literature studies. He also studied few semesters in physics and computer science. He currently works as a research software engineer at University of Münster, Germany.

## Notes

---

1. <https://icu.unicode.org/home>. ↩
2. <https://github.com/SCDH/icu-xpath-bindings>, implemented by one of the authors of this paper. ↩
3. First stylesheets for transliterating documents and ODDs are in the repository of this paper: <https://github.com/lueck/tei2023>. ↩

# 3D Text Encoding and TEI: Text, Editions, and Spatiality

Jun Ogawa, ROIS-DS Center for Open Data in the Humanities, Japan

Kiyonori Nagasaki, International Institute for Digital Humanities

Asanobu Kitamoto, ROIS-DS Center for Open Data in the Humanities, Japan

Poster

**Keywords:** TEI, 3D, sourceDoc, Linked Data, visualization

## Abstract

The incorporation of 3D technologies within the realm of digital humanities has emerged as a captivating subject for numerous scholars. A significant challenge lies in determining the most effective representation of interconnected information, encompassing meta- and para-data of 3D models, images, movies, and textual content, within this innovative multi-dimensional environment<sup>[1]</sup>.

To address this, we designed a conceptual model for a 3D knowledge network (Fig. 1), ensuring more accurate utilization of 3D information in terms of academic credibility, multiple interpretations, and data usability. While preceding studies like ADS Data Model<sup>[2]</sup>, SCOTCH<sup>[3]</sup>, or HBIM<sup>[4]</sup> have touched on similar discussions for 3D information description, the specific challenge of 3D text representation and the interplay between textuality and spatiality remain insufficiently explored even in those works. How can we effectively encode text inscribed on 3D objects or positioned within 3D space, preserving its spatial context while maintaining a well-structured text edition? This crucial aspect warrants further investigation to enhance the integration of textual content within the 3D landscape.

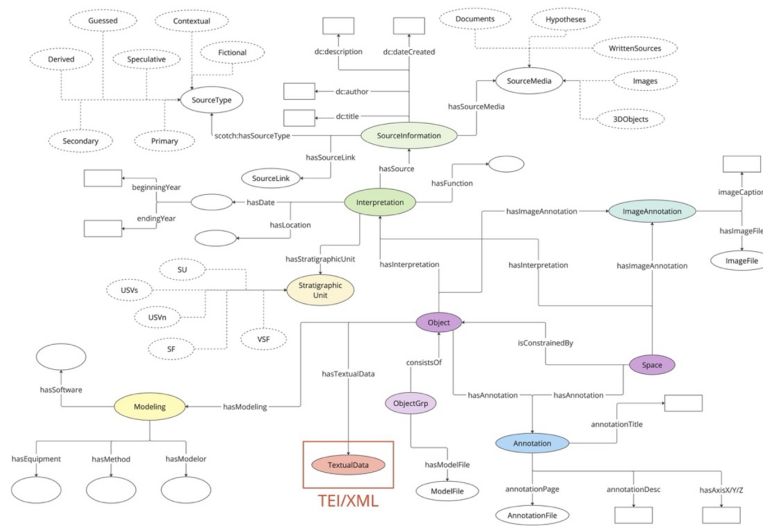


Fig. 1: Conceptual model for the 3D information

In this poster, we present an enhanced method for representing text in a 3D context, building upon our previous discussions at TEI 2022 where we suggested the use of the `<sourceDoc>` [5]. Our newly developed approach allows for a more refined mapping of text data and 3D spatial information, down to the character level, by leveraging the `<c>`. Additionally, we have extended this method to automatically render detailed textual features encoded in existing TEI tag sets, such as `<corr>`, `<expan>`, or `<del>`, into the 3D space. Through these advancements, we achieve a more practical and elaborate representation, bridging the gap between textual content and the 3D environment effectively.

```
<sourceDoc rend="3D">
  <surfaceGrp xml:id="inscription1" type="3Dobject">
    <figure>
      <graphic url="--URI of a 3D object--" />
    </figure>
    <surface xml:id="sf_00" type="position">
      <zone xml:id="z_01_sf_00" ulx="-0.02" uly="5.51" ana="z:-0.27" rotX="0" rotY="0" rot-Z="0" />
      <zone xml:id="z_05_sf_00" ulx="-0.25" uly="5.51" ana="z:-0.3" rotX="0" rotY="0" rot-Z="0" />
    </surface>
    <surface xml:id="sf_01" type="text" ana="edition" xml:lang="lat" corresp="#wit_1 #wit_2">
      <lb n="1" xml:id="z_01_sf_01" facs="#z_01_sf_00"/>
      <expan>
        <abbr>
          <c xml:id="c_0001" facs="#z_05_sf_00">L</c>
        </abbr>
        <ex>ucio</ex>
      </expan>
    </surface>
  </surfaceGrp>
</sourceDoc>
```

Fig. 2: Example of 3D text encoding

Fig. 2 represents a fragment of sample EpiDoc markup, focusing on a single word within a specific inscription: "L(ucio)." The word "L(ucio)" is unequivocally identified as an ablative form of a Latin personal name, and it is accompanied by an

abbreviation. This abbreviation is systematically encoded using the `<expan>`, following the specification of the EpiDoc schema. Interestingly, a discrepancy emerges between our model and the EpiDoc schema concerning the placement of the aforementioned text. While the EpiDoc schema traditionally locates such text within the `<text>` block, our model takes a different approach. In our proposed framework, the text is encoded within the `<surface>`, incorporating an attribute of `@type="text"`, which serves as a grandchild element of the `<sourceDoc>`.

To identify the letter 'L' within the text, we utilize a `<c>` tag, augmented with an `@fac` value corresponding to the `<zone>` in another `<surface>` block with `@type="position"`. This strategic linkage enables the explicit representation of spatial information, encompassing X-Y-Z coordinate and rotation. As a consequence, the encoded text can be seamlessly rendered into the 3D space. As our model also permits the connection of lines to the `<zone>` information, entire lines within the inscription can be displayed in the 3D realm.

In our research, we not only propose novel encoding methods for representing 3D textuality but also take a step further by transforming TEI-encoded data into a Linked Open Data (LOD) resource. By doing so, we establish valuable connections within a vast network of 3D knowledge, transcending the boundaries of individual texts or objects. The concept of this interconnected network is visually depicted in Fig. 1. Specifically, we model TEI/XML data as an instance of the `:TextualData` class, showcasing its importance in preserving textual information. This textual data is then linked to an `:Object` instance through the `:hasTextualData` property, establishing a meaningful relationship between the represented object and its associated textual content.

Besides suggesting the new encoding methods for 3D textuality, we decided to make a TEI-encoded data function as a LOD resource, linking it to the wider network of 3D knowledge not limited to any single text or object. Such a network has already been shown in Fig. 1, where TEI/XML data is represented as an instance of `:TextualData` class connected to an `:Object` instance by `:hasTextualData` property. We have developed a 3D visualization system (Fig. 3) that enables the visualization of such a knowledge cloud, capturing the intricate connections and relationships within a vast repository of knowledge.



Fig. 3: A 3D viewer displaying the text encoded in TEI/XML, Transcription cited from I.L.N. [6]

Through this approach, we bridge the gap between textual information and 3D representations, enabling seamless integration and accessibility of data within the broader 3D knowledge landscape. This interconnectedness not only enhances the understanding of individual texts and objects but also fosters a more comprehensive and interconnected understanding of 3D textuality as a whole.

## Notes

1. Susan Schreibman & Costas Papadopoulos (2019), 'Textuality in 3D: three-dimensional (re)constructions as digital scholarly editions,' *International Journal of Digital Humanities*, vol. 1, pp. 221-233. [↩](#)
2. Martina Trognitz, Kieron Niven and Valentijn Gilissen (2016), "3D Models in Archaeology: A Guide to Good Practice." In *Guides to Good Practice*, by Archaeology Data Service / Digital Antiquity. University of York, UK: Archaeological Data Service. [↩](#)
3. Valeria Vitale (2017), 'Rethinking 3D digital visualization: from photorealistic visual aid to multivocal environment to study and communicate cultural heritage,' PhD Thesis, King's College London. [↩](#)
4. Martina Attenni (2019), 'Informative Models for Architectural Heritage,' *Heritage* 2019, pp. 2067-2089. [↩](#)
5. Jun Ogawa et al. (2022), 'Poster: Text as Object: Encoding the data for 3D annotation in TEI,' *TEI 2022 Conference Book*, pp. 86-88. [↩](#)

6. Bernard Rémy et al (2004), *Inscriptions Latines de Narbonnaise (I.L.N.)*, V. 2.  
Vienne, CNRS Editions, p. 268. [↩](#)



# A (cautionary) tale of two texts

**Nick Thieberger**, University of Melbourne, Australia

**Robert Turnbull**, University of Melbourne, Australia

**Daniel Russo-Batterham**, University of Melbourne, Australia

**Birgit Lang**, University of Melbourne, Australia

Short Paper

**Keywords:** TEI, HTML, Encoding

## Abstract

The Text Encoding Initiative (TEI) community is replete with wonderful encoded documents and images of manuscripts. Given the number of such projects and the effort that has gone into the TEI itself, there are many frameworks, tools, and workflows that can be used to meaningfully encode and digitally present manuscript sources. In this presentation, we reflect on our experiences working on two contrasting manuscripts in an institutional environment where TEI has little uptake. In particular, we explore some of the challenges and tradeoffs we encountered creating digital editions with only limited institutional support for sustainable Digital Humanities research software infrastructure and training.

We are humanities technologists (DR, RT) and humanities researchers (BL, NT) at the University of Melbourne in Australia. The first manuscript we worked with is a handwritten German text (BL) of some 100,000 words in 391 pages, to which we added a transcript, notes, facsimiles, and a translation. We used the Text Encoding Initiative to richly capture a German and English version of this manuscript, encoding in the TEI people, places, bibliographical references, and fictional characters. This was published online using TEI Publisher Web Components and required the team (DR, RT) to create a virtual machine, build a Django interface, provision a IIIF server, provision storage for images, and maintain the site over time, all of which incurs significant technical debt and requires specialised skills.

The second manuscript (NT) comprises 8,000 words of ethnographic notes from Vanuatu in 1914. Multiple versions of the same text were combined to make a single diplomatic edition that allows a reader to follow the content in a way that was not possible with the original documents. An HTML version built by NT presents the text and images of the manuscript originals, sometimes up to eleven different page images corresponding to the same text, with decisions required to arrive at a

consensus document. It is housed on a site controlled by NT, and is picked up by the Internet Archive, as it is a simple HTML page. It requires no maintenance and has no dependencies, and NT was able to build the site himself.

We will discuss some of the following questions based on our experience at the University of Melbourne. How can we scope projects, understand the workload implications, and manage the expectations of academics who become excited after seeing completed TEI projects and want to apply the technology in their work? How can non-technical Humanities scholars use these technologies, and how can they ensure longevity of their work beyond their interaction with technical colleagues? What kind of ongoing support is required to keep a site like this going? This includes basic necessities like a domain name and site, but also encompasses the dependencies of the selected tools.

While some institutions have TEI support services that can guarantee ongoing access to encoded texts, what is the best strategy for an academic who does not have access to local TEI support services that can guarantee ongoing access to encoded texts?

# A digital annotated parallel corpus of Athonite text types: Gospel of Matthew

Eka Kvirkvelia, University of Hamburg, Germany  
Tamar Kalkhitashvili, Ilia State University, Georgia

Long Paper

**Keywords:** Digital edition, Georgian translation of the four Gospels, TEI (Text Encoding Initiative)

## Abstract

The paper aims to present the results of the ongoing research project *“Digital edition of an annotated parallel corpus of Giorgi Athonite’s recension of the Gospel of Matthew”* (2021-2023) funded by Shota Rustaveli National Science Foundation (YS-21-1562). The project is carried out on the technical base of the host institution: Ilia State University (Tbilisi, Georgia).

The translation of the New Testament and generally biblical texts into Georgian is supposed to have been made since the 4th-5th centuries. The Georgian translation of the four Gospels takes into consideration different traditions. Hence, the manuscripts of the four Gospels that have come down to us preserve non-homogeneous, diverse texts, namely: **Pre-Athonite** text types (Adishi, Opiza (the same Jrutchi-Parkali) and mixed); **Athonite** text types (so-called Ekvtimiseuli (belonging to Euthymius) and Giorgi Athonite’s recension) and the **Hellenophilic** text type.

Within the project, we are in the process of preparing a parallel digital edition of **Athonite** text types of Matthew Gospel: a) the final recension of George the Athonite (1009-1065); b) the text reflecting the intermediate stage of George the Athonite’s work; c) and the text of the Gospel preserved in John Chrysostom’s Commentary of the Gospel of Matthew translated by Euthymius the Athonite (955-1028), as well as the English and Greek (Byzantine type) texts. Among the Athonite text types Giorgi the Athonite’s final recension has the most significance, as it is established in the Georgian Church up to these days. Due to this reason, within the current project this text was chosen for annotation. A digital annotated parallel

corpus of Athonite text types is based on the theoretical research provided within my doctoral thesis.<sup>[1]</sup>

The digital edition of Georgian medieval manuscripts based on the TEI has been an ongoing project at Ilia State University since 2009. Dr. Irina Lobzhanidze has been instrumental in digitizing Georgian medieval manuscripts, particularly the Typicon of the Georgian Monastery of the Holy Cross Near Jerusalem, which was published using eXist-db and TEI Publisher (Lobzhanidze 2021)<sup>[2]</sup>, employing IIIf manifests of the manuscripts.

The technical team of the current project experimented with exist-DB, but due to limited time and funds from a young researcher grant, they decided to create a flexible platform with a graphical interface for annotating texts. The main goal was to reflect the findings of the doctoral research and publish texts online. The platform (<http://ogb.iliauni.edu.ge/>) features a user-friendly annotation tool built using standard HTML, CSS, and Javascript, connected to a PostgreSQL database. Researchers can annotate specific manuscript texts, with updates simultaneously reflected in the database and displayed on the OGB webpage alongside parallel annotated texts. Hovering over the texts reveals the annotation records. The platform also allows users to download annotated texts in XML format, with an eye on future collaboration for reusing TEI XMLs of Georgian medieval texts and creating a TEI schema.

Here is the result of the annotation for one of the files:

```
</teiHeader>
  <text>
    <body>
      <div type="edition" xml:space="preserve">
        <div type="chapter" xml:id="cha.1">
          <l n="cha.1.1">
            <seg xml:lang="ka">
              ნიკნი შობისა იესუ ქრისტესი, ძისა დავითისი,
              ძისა აბრაჰამისი.
            </seg>
            <seg xml:lang="grc">
              Βίβλος γενέσεως Ἰησοῦ χριστοῦ,
              υἱοῦ Δαυίδ, υἱοῦ Ἀβραάμ.
            </seg>
          </l>
        </div>
      </div>
    </body>
  </text>
</teiHeader>
```

```

<entry xml:id="w.0.cha.1.1">
  <form type="lemma">ბიბლია
    <lbl xml:lang="grc">βίβλος</lbl>
    <lbl xml:lang="arm">Գիրք</lbl>
    <lbl xml:lang="en">book</lbl>
  </form>
  <gramGrp>
    <gram type="pos">N.Nom.Sg</gram>
  </gramGrp>
</entry>
<entry xml:id="w.1.cha.1.1">
  <form type="lemma">Ծնունդ
    <lbl xml:lang="grc">γενέσεως</lbl>
    <lbl xml:lang="arm">ծննդեան</lbl>
    <lbl xml:lang="en">birth</lbl>
  </form>
  <gramGrp>
    <gram type="pos">N.Gen.Sg</gram>
  </gramGrp>
</entry>

```

The aim of the project is to create a high-quality digital edition that will serve as a model for the digital publication of other Gospels (of Mark, Luke, and John), as well as other books of the New and Old Testaments based on textological research. Regretfully, however, we do not yet have either a Scholarly edition of the Georgian translations of the four Gospels or an annotated parallel corpus of a digital edition, due to which various text types of Georgian translations of the four Gospels remain inaccessible to foreign scholars. For this reason, the Georgian text is not included in any digital editions of the Greek Gospels and their translations. In such circumstances, an annotated parallel corpus of the Gospel of Matthew has special importance for further research.

## About the authors

**Dr Eka Kvirvelia** – Research Associate at the Centre for the Study of Manuscript Cultures (CSMC), the University of Hamburg, team member of the project “*The Development of Literacy in the Caucasian Territories (DeLiCaTe)*” led by Prof. Dr. Jost Gippert. <https://www.csmc.uni-hamburg.de/research/current-projects/delicate.html>

Eka Kvirkvelia defended a PhD thesis titled "*Codicological and Textological Analysis of 11th-Century Georgian Manuscripts of the Gospel and the History of Formation of Giorgi Athonite's Recension of the Four Gospels*" at Tbilisi State University (Georgia). The research aimed to conduct a palaeographical, codicological, and textual analysis of 11th-century manuscripts containing the Georgian translation of the Gospels which reflect the editorial work of George the Athonite.

Eka Kvirkvelia is a head of the ongoing research project for young scientists "*Digital edition of an annotated parallel corpus of Giorgi Athonite's recension of the Gospel of Matthew*" (2021-2023), funded by Shota Rustaveli National Science Foundation (Georgia). The project is carried out on the technical base of the host institution: Ilia State University (Tbilisi, Georgia), under the guidance of Prof. Nino Doborjginidze, and by consulting Prof. Darejan Tvaltvadze (Ivane Javakhishvili Tbilisi State University).

Eka Kvirkvelia was a team member of the Project "*The Epigraphic Corpus of Georgia (Standardized digital edition)*" (2017-2020), which implied preparing a standardized bilingual digital edition of the rich epigraphic evidence discovered or preserved in Georgia (Georgian, Greek, Armenian, Aramaic, Hebrew), encoding inscriptions according to the standards of EpiDoc. Her responsibility in this project was encoding the Georgian inscriptions, and preparing XML files in Oxygen XML Editor. For the online publication of inscription's texts is used EpiDoc Front-End Services (EFES) platform. <https://epigraphy.iliauni.edu.ge/en/>

Eka Kvirkvelia's research areas are Georgian Palaeography and Codicology, Old Georgian Language, History of the Georgian Literary Language, Georgian Epigraphy, Digital Epigraphy, Textual criticism of the Georgian Four Gospels, Digital edition of biblical texts, Georgian Hymnography.

**Dr Tamar Kalkhitashvili** – Researcher at The Institute of Linguistic Studies of Ilia State University.

Tamar Kalkhitashvili was the coordinator of the "*The Epigraphic Corpus of Georgia (Standardized digital edition)*" (2017-2020) funded by Shota Rustaveli National Science Foundation, her main responsibility, besides managing the project, was editing XMLs of multilingual inscriptions according to EpiDoc guidelines and building the corpus in EFES.

Tamar Kalkhitashvili was a team member of the projects "*Multilingual Parallel Corpus of the Book of Tobit*" ([books.tobit.iliauni.edu.ge](https://books.tobit.iliauni.edu.ge)), the paper "Multilingual

Annotated Electronic Parallel Corpus-based Edition of the Georgian translation of the Book of Tobit" has been accepted for publishing in the journal *Le Muséon* (2023)) and "*Digital edition of an annotated parallel corpus of Giorgi Athonite's recension of the Gospel of Matthew*" (<http://ogb.iliauni.edu.ge/>), where she was responsible for coordinating the technical team.

Main research interest areas of Tamar Kalkhitashvili are Digital Humanities, Digital Epigraphy, Corpus Linguistics and digital publication of literary and historical texts in underdescribed languages.

## Notes

---

1. Eka Kvirkvelia defended the doctoral thesis titled "*Codicological and Textological Analysis of 11th-Century Georgian Manuscripts of the Gospel and the History of Formation of George the Athonite's Recension of the Four Gospels*" for Ph.D. in Philology at the Ivane Javakhishvili Tbilisi State University (Tbilisi, 2019), link: [http://press.tsu.ge/data/image\\_db\\_innova/EkaKvirkvelia.pdf](http://press.tsu.ge/data/image_db_innova/EkaKvirkvelia.pdf). ↵
2. Irina Lobzhanidze. 2021, *Georgian Medieval Manuscripts and the Digital Edition of the Typicon of the Georgian Monastery of the Holy Cross near Jerusalem*. In: *Le Muséon*: 134, pp. 325-352 (3-4). ↵

# A multi-media dictionary of endangered languages with TEI Lex-O: a case study of Hatoma, Yaeyama Ryukyuan

**Natsuko Nakagawa**, National Institute for Japanese Language and Linguistics, Japan

**So Miyagawa**, National Institute for Japanese Language and Linguistics, Japan

**Poster**

**Keywords:** TEI Lex-O, endangered language, lexicography, Hatoma, Ryukyuan

## Abstract

Recently, the authors and their colleagues have published several digital as well as printed dictionaries of Ryukyuan languages spoken in Okinawa and Kagoshima Prefectures, Japan. Ryukyuan languages are endangered, as only the grandparent generation speaks them in everyday conversation. Younger generations mainly use Japanese; some of them understand but do not use the languages, while others neither understand nor use them (Hanmine 2020). Searchable digital dictionaries with audio information help potential Ryukyuan speakers to learn their local languages.

Currently, we are building a digital dictionary of Hatoma, a Yaeyama Ryukyuan language, in the TEI Lex-O format (Romary and Tasovac 2018). The original dictionary was written by Shinichi Kajiku, a native speaker of Hatoma Ryukyuan and a linguist. He spent more than 50 years compiling the dictionary, which was structured in a spreadsheet format by one of the authors in 2019 (Kajiku and Nakagawa 2020). The structured dictionary with audio files for each entry is now available online.

The dictionary contains some elements that TEI Lex-O does not assume yet. For example, the guideline does not mention how to encode audio files. We decided to use the media element to provide the information on audio files. Additionally, the guideline does not provide elements related to tones and accents, while it does provide the stress element. We decided to use the `<pron>` element with the attribute `@notation="accent"` to provide accent information in Hatoma



Ryukyuan. Finally, the original dictionary provides information on the pronunciation of example sentences in addition to entries. The example sentences and entries are written in kana, a Japanese writing system, which is a reader-friendly writing system but does not perfectly provide information on how to pronounce each expression. We decided to use the <pron> element with @notation="ipa" for example sentences although it is currently meant for only entries.

## Bibliography

Kajiku, Shinichi, and Natsuko Nakagawa. (2020). *Hatoma-Japanese Dictionary*. Tachikawa: NINJAL Language Variation Division.

Hanmine, Madoka. (2020). *Speaking my Language and Being Beautiful – Decolonizing Indigenous Language Education in the Ryukyus with a Special Reference to Sámi Language Revitalization*. Rovaniemi: Lapin Yliopisto.

Romary, Laurent and Toma Tasovac. (2018). "TEI Lex-0: A Target Format for TEI-Encoded Dictionaries and Lexical Resources." TEI 2018 Conference.

## About the authors

**Natsuko Nakagawa** is an associate professor at the National Institute for Japanese Language and Linguistics. She is a linguist doing fieldwork on Yaeyama Ryukyuan and Nambu dialect in Aomori. Her research interest is how people within and across the research areas can communicate with each other using digital data.

**So Miyagawa** is an assistant professor at the National Institute for Japanese Language and Linguistics, working on digitization projects of language resources (text, audio, video, image) of endangered Ryukyuan languages. His research interests are in cultural heritage preservation and revitalization.

# A Preliminary Proposal for Digital Scholarly Editing that Uses Modern Japanese autograph manuscripts: How to markup autograph manuscripts of Rampo Edogawa

Sachiko Shioi, Faculty of Letters, Arts and Sciences, Graduate school of Waseda University, Japan

Kiyonori Nagasaki, International Institute for Digital Humanities

Poster

**Keywords:** manuscript, digital scholarly editing, genetic editing, transcription

## Abstract

Recently, digital scholarly editing has been discussed in the research on modern autograph manuscripts<sup>[1]</sup>, such as archives in which documentary texts are the main focus of editing<sup>[2]</sup>, and attempts to enable an intuitive understanding of the process of writing and creation<sup>[3]</sup>. On the other hand, in the research on modern Japanese autograph manuscripts, digital scholarly editing has yet to be conducted because, in Japan, digital images and paper publications with transcriptions and images are used as surrogates of autograph manuscripts. We have already discussed the application of the TEI guidelines to modern Japanese autograph manuscripts<sup>[4] [5]</sup>. Therefore, we discuss digital scholarly editing that uses TEI, pointing out research questions on how to provide its surrogates in Japan.

Three questions currently exist in the research on modern Japanese autograph manuscripts in Japan: usability, representation of variants, and transcription. In modern Japanese autograph manuscripts, authors used kuzushi-ji (figure 1), which profoundly differs from contemporary Japanese handwriting and is difficult even for modern Japanese researchers to read.

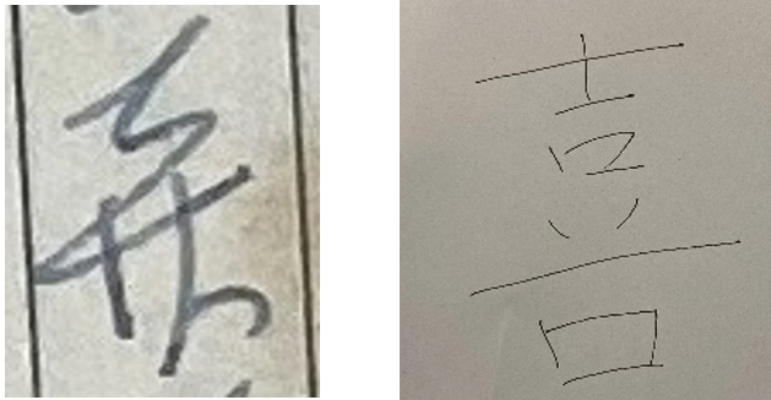


Fig. 1: On the left is the kuzushi-ji form of 喜 from *Nisen-doka* manuscript, on the right is the contemporary Japanese handwriting form of 喜.

Thus, it is inconvenient to use them for research if only digital images are provided. Furthermore, paper publications with transcriptions and images are difficult to present variants within a manuscript since only the "final form" becomes the text. When indicating the writing process, it becomes complicated and hard to understand. In addition, in some cases, it is impossible to represent the scribal scripts of kuzushi-ji in the paper edition, and accuracy is sacrificed for the sake of readability. Because of these questions, research on modern manuscripts is unfavorable in Japan.

To resolve these questions, we propose to markup autograph manuscripts of the works by Edogawa Rampo, one of Japan's earliest writers of detective stories, based on a document-oriented markup method. In his works it is often observed that he writes on the blank pages of his diary, which is atypical considering that writers of modern Japanese literature typically use manuscript paper. However, given the complexity of his writing process and the abundance of examples utilizing cursive script, this case is deemed appropriate for study.

## About the authors

**Sachiko Shioi** is a Ph.D. student at Faculty of Letters, Arts and Sciences, Graduate school of Waseda University. She researches autograph manuscripts of works by Edogawa Rampo, one of Japan's earliest writers of detective fiction. Her project is a digital scholarly editing that uses modern Japanese autograph manuscripts.

**Kiyonori Nagasaki**, Ph.D., is a senior fellow at the International Institute for Digital Humanities in Tokyo. His main research interest is developing digital frameworks for

collaboration in Buddhist studies. He is also investigating the significance of digital methodology in the Humanities and promoting DH activities in Japan.

## Notes

---

1. Driscoll, M. J., Pierazzo, E. (Eds.). (2016). *Digital scholarly editing: Theories and practices*. Open Book Publishers, pp.49-50. [↵](#)
2. Cohen, M., Folsom, E., Price, M. K., Editors. (n.d.). The Walt Whitman Archive <https://whitmanarchive.org/> (last accessed: 29/04/2023). [↵](#)
3. Pierazzo, E., André, J. (n.d.). "Autour d'une séquence et des notes du Cahier 46: enjeu du codage dans les brouillons de Proust": [http://peterstokes.org/elena/proust\\_prototype/](http://peterstokes.org/elena/proust_prototype/) (last accessed: 25/04/2023). [↵](#)
4. Shioi, S., Nagasaki, K. (2022). "A proposal of an XML markup method based on TEI using a 'Nisen-dōka' manuscript by Rampo Edogawa". *IPSJ SIG Technical Report*, Vol. 2022-CH-130 No.4. [↵](#)
5. Shioi, S., Nagasaki, K. (2022). "Possibility of Structural Description for handwritten materials of modern Japanese literature: Using Rampo Edogawa's handwritten materials". *IPSJ Symposium Series* Vol. 2022, No.1. [↵](#)

# “Ain schone kunstliche Underweisung”: Modelling German lute tablature in MEI

Reinier de Valk, Moodagent A/S, Copenhagen

Kateryna Schöning, Universität Wien

David Weigl, Universität für Musik und darstellende Kunst Wien

David Lewis, University of Oxford e-Research Centre

Tim Crawford, Goldsmiths University of London, United Kingdom

Marc Lewon, Schola Cantorum Basiliensis

Paul Overell, Independent Scholar

Long Paper

**Keywords:** German lute tablature, instrumental tablature, MEI modelling

## Abstract

Among the lute tablature systems that were in use during the Western lute’s lifespan (ca. 1450-1800), German lute tablature, which was only used in German-speaking countries and fell out of fashion after 1600, takes a special place. It stands apart from the other main systems—French, Italian, and Spanish lute (or vihuela) tablature—in that it does not provide a visual representation of the instrument. That is: rather than using a set of letters or numbers (indicating frets) on horizontal lines (indicating courses), showing the lutenist where to place their fingers on the fretboard and which courses to pluck, German lute tablature ‘encodes’ all course-fret intersections with unique symbols (letters, numbers, and other). This renders the notation more abstract and less intuitive – even for lutenists.

While modelling the ‘staff-based’ French, Italian, and Spanish lute tablature systems in MEI is relatively straightforward (work on this is ongoing and already in advanced state), modelling German lute tablature brings along a number of specific challenges. In collaboration with the MEI Tablature Interest Group, an MEI model for German lute tablature is currently being developed. This is done as part of E-LAUTE, an international Weave (FWF/DFG/SNSF)-funded project that aims to investigate and unlock a corpus of music that is almost exclusively notated in German lute tablature, and for which such a model is indispensable.

In this paper, we report on the progress on this task, discussing the specific challenges we encounter, as well as our proposed solutions. Among these challenges are (i) the multitude of and variety in symbols used in the different sources; (ii) the absence of a staff, and, related to this, the issue of vertical positioning of the symbols; (iii) further development of the music notation engraving library Verovio to support and render German lute tablature, and the issues associated with this; (iv) automatic conversion of existing file formats for encoding (German) lute tablature into MEI, and how existing tools (Fronimo, luteconv, mei-friend) can be used in this process; and (v) the similarity of German lute tablature to keyboard tablatures, a substantial repertory, for which as of yet no MEI model exists, but whose implementation must be anticipated.

# Alaine Locke's *New Negro Anthology* in TEI and MEI: A DEFCON Teaching Collaboration across Two Universities

Elisa Eileen Beshero-Bondar, Penn State Erie, The Behrend College, United States of America

Bartholomew Brinkman, Framingham State University, United States of America

Short Paper

Keywords: digital humanities, pedagogy

## Abstract

In 2021-2022 the Digital Ethnic Futures Consortium (DEFCON) organized a mentorship program for professors to enhance their teaching of Digital Humanities.<sup>[1]</sup> The coauthors are a DEFCON mentor and mentee whose regular conversations inspired pedagogical experiments. One of us specializes in teaching an undergraduate DH curriculum, while the other is introducing DH in the context of a literature course on the Harlem Renaissance. We each recognized that our teaching lacked something. The literature professor wanted to introduce students to TEI but did not have more than two or three weeks in a semester to attempt it. The coding professor had a semester to teach markup technologies but lacked time to explore poetry intensively. What if we could supply something the other lacked?

Four students from the DH coding class formed a project team that agreed to work for the co-author's Harlem Renaissance literature class as their "clients" to create a new digital resource from Alaine Locke's *New Negro Anthology*.<sup>[2]</sup> We wanted the literature students to learn XML markup and see their markup visualized in an interactive website. We sought for the coding students to gain awareness of the complexities of poetic language and music, even as they programmed XSLT to make a static HTML-based web edition. The coding team learned MEI to digitize short bars of music in the anthology and make them playable on their website. Hosted freely on GitHub (<https://newtfire.github.io/locke-anthology/>), this digital teaching edition has room to grow and continue engaging students with poetry, music, and code.

## Notes

---

1. The coauthors wish to thank Roopika Risam and the DEFCON steering committee for organizing the generous collaboration and exchange opportunities of the DEFCON mentoring program. For more information see <http://digitaletnicfutures.org/>. ↵
2. first published in 1925 and now in the US public domain. We started with a facsimile shared on the Internet Archive (<https://archive.org/details/newnegrointerpre00unse/page/156/mode/2up>). ↵



# An AI-assisted Digital Scholarly Editing System for Buddhist Studies.

Kiyonori Nagasaki, International Institute for Digital Humanities  
Masahiro Shimoda, The University of Tokyo

Poster

**Keywords:** critical apparatus, Buddhist studies, digital scholarly edition, double-endpoint-attached method, deep-learning OCR

## Abstract

In recent years, IIIF-compatible images of several xylographic Daizokyos (DZK, Series of Buddhist scriptures) have become available, mainly from Japanese research institutions and temples. Each DZK consists of approximately 5,000 to 6,000 scrolls converted into 150,000 to 190,000 digital images, and some used as a witness for the modern letterpress DZK and others not. In addition, printed materials that are fragments of them, as well as manuscripts that precede them, have been made available with IIIF by cultural institutions around the world. The situation has made authors build a Buddhist text database (Nagasaki, Tomabechi & Shimoda 2013) aggregating the IIIF images from various institutions. It enables researchers to have an opportunity to re-edit the existing critical edition, which is sometimes noted as needing re-editing due to errors or oversights. However, although the current system has made it possible to compare the page images of the witnesses with each other, the only way to check variants in the texts on the witnesses is to check each letter manually, which was time-consuming.

One way to solve this problem is to automatically contrast an OCRed text with the digital transcriptions of the existing edition if highly accurate OCR is available. However, until last year, although the development of OCR has long been undertaken, it has been challenging to find a suitable solution because many of the documents are manuscripts or woodblocks, which are not stable in character form, and because there are many different types of characters.

Under these circumstances, last year, the Next Generation Lab at the National Diet Library in Japan released an open-source OCR software (NDL OCR 2023) with

deep learning technology for East Asian classics, which is more accurate than the previous ones, enabling automatic comparison between each OCR'd diplomatic text (Text-D) and our text data digitized from a critical edition (Text-CE). The system detects variants between Text-D and Text-CE using the difflib module in Python.

Users of the system see the place of variants as yellow-marked texts on the windows of Text-CE (left window) and Text-D (center window). Users can see the variants on text data and an area shaded in red on the OCR'd digital facsimile in IIF, including the place of the variant characters. In Fig.1, a variant noted in the footnote of the critical edition has been confirmed via the OCR'd text with the corresponding IIF image. The system also allows users to find errors or oversights similarly. That is, if a yellow-marked character is different between both texts, a variant is not noted in the footnote, and the accuracy of the character in the OCR'd text is confirmed with the IIF image, it should be an oversight in the existing edition. Occasionally, textual errors may occur due to misidentification by the OCR software. Even in such cases, automatic contrast with Text-CE can significantly increase the likelihood of detecting errors.

Moreover, determining whether or not a mistake has been made is also easy to perform, as the IIF image of the corresponding part can be easily focused with a single click. After checking the texts, authorized users can revise such errors in both texts in the system. As our project adopts TEI guidelines with double-endpoint-attached method, the system can include various witnesses without overlapping.

This means that the scholarly editing system using OCR text has finally become a realistic possibility in East Asian Buddhist studies.

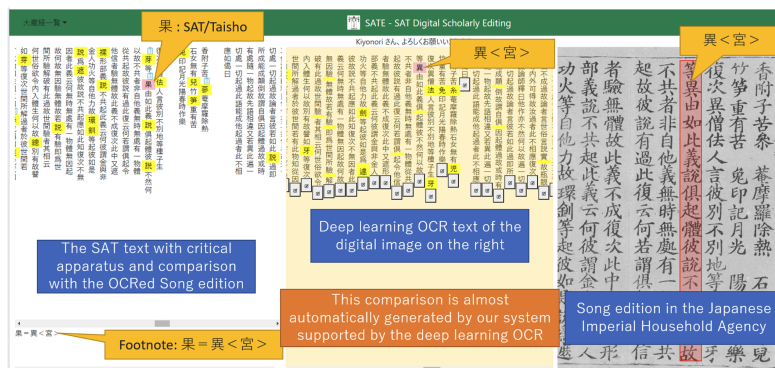


Fig. 1: A typical interface of the system

## Bibliography

Nagasaki, K., Tomabechi, T., & Shimoda, M. (2013). Towards a Digital Research Environment for Buddhist studies. *Literary and Linguistic Computing*, 28(2), 296–300. <https://doi.org/10.1093/lc/fqs076>

NDL OCR application for classics (NDL 古典籍OCRアプリケーション), [https://github.com/ndl-lab/ndlkotenocr\\_cli](https://github.com/ndl-lab/ndlkotenocr_cli), accessed 2023-04-30.

## About the authors

**Kiyonori Nagasaki**, Ph.D., is a senior fellow at the International Institute for Digital Humanities in Tokyo. His main research interest is developing digital frameworks for collaboration in Buddhist studies. He is also investigating the significance of digital methodology in the Humanities and promoting DH activities in Japan.

**Masahiro Shimoda**, Ph.D., is a former professor at the University of Tokyo, and a professor at Musashino University. He studies the formation process of Mahayana Buddhist scriptures and is also leading a project to build and develop the database of Buddhist Chinese texts (SAT).

# An annotation model for software mentions and citations

**Fernanda Alvares Freire**, Universität Rostock, Germany

**Anne Ferger**, University of Paderborn, Germany

**Ulrike Henny-Krahmer**, Universität Rostock, Germany

**Daniel Jettka**, Universität Paderborn, Germany

**Short Paper**

**Keywords:** research software, software citation, annotation, TEI taxonomy

## Abstract

Appropriate citation of software plays an important role in academic publications to make research results reproducible and reusable. There are several recommendations and guidelines on how to deal with research software (Anzt et al. 2021; Smith et al. 2016; Lamprecht et al. 2020) and how to cite software that is used in the research process (Jackson n.d.; Chue Hong et al. 2019a, 2019b; Druskat 2021a, 2021b).

To find out if these recommendations (e.g. consistent versioning, persistent identification, appropriate credit to developers) are actually reflected in practice, we examined conference abstracts of the DHd (Henny-Krahmer/Jettka 2022) and ADHO conferences (Jettka et al., to appear). Apparently, there is great potential (and need) for improving the current situation. In our software citation studies, an annotation model was formulated in the form of a TEI taxonomy. Initially, a document-centered approach was pursued, i.e., software mentions were semi-automatically identified and directly annotated with citation information that could be located somewhere else in the document, for instance, in the bibliographies of the abstracts.

We now propose a revised TEI annotation model, which aims at a more precise annotation to differentiate between pure software mentions (names) and other parts of citation information (such as URLs, developers, or bibliography entries) and linking these parts together. In our new approach, we use pointers (<ptr>) to an externally defined list of software entities and reference these pointers from the annotation instances of corresponding citation information. Thus we still aim at

examining the current situation of software citation (this time in articles of the Journal of the Text Encoding Initiative), but at the same time, we provide and use a model and create a data basis of software mentions and citations which could be used for training of automatic methods to identify software mentions and corresponding citation information in academic texts.

## Bibliography

Anzt, Hartwig / Bach, Felix / Druskat, Stephan / Löffler, Frank / Loewe, Axel / Renard, Bernhard Y. / Seemann, Gunnar / Struck, Alexander et al. (2021): "An environment for sustainable research software in Germany and beyond: current state, open challenges, and call for action [version 2; peer review: 2 approved]." In: *F1000Research* 9:295. <https://doi.org/10.12688/f1000research.23224.2>.

Chue Hong, Neil (ed.) (2019a): "Software Citation Checklist for Authors" (Version 0.9.0). *Zenodo*. <http://doi.org/10.5281/zenodo.3479199>.

Chue Hong, Neil (ed.) (2019b): "Software Citation Checklist for Developers" (Version 0.9.0). *Zenodo*. <http://doi.org/10.5281/zenodo.3482769>.

Druskat, Stephan (2021a): "Research software citation for researchers." *Research Software Citation. Cite and Make Citable!* (Version 1.1). <https://cite.research-software.org/researchers/> [last accessed: 04.11.2022].

Druskat, Stephan (2021b): "Research software citation for developers." *Research Software Citation. Cite and Make Citable!* (Version 1.1). <https://cite.research-software.org/developers> [last accessed: 04.11.2022].

Henny-Krahmer, Ulrike / Daniel Jettka (2022): "Softwarezitation als Technik der Wissenschaftskultur. Vom Umgang mit Forschungssoftware in den Digital Humanities." In: *DHd 2022. Kulturen des digitalen Gedächtnisses. Konferenzabstracts, Potsdam, 7.-11.3.2022*. Potsdam: Universität Potsdam, 203-206. DOI: <https://doi.org/10.5281/zenodo.6328047>.

Jackson, Mike (n. d.): "How to cite and describe software." In: *Software and research: The Software Sustainability Institute's Blog*. <https://www.software.ac.uk/how-cite-and-describe-software> [last accessed: 04.11.2021].

Jettka, Daniel / Henny-Krahmer, Ulrike / Ferger, Anne / Alvares Freire, Fernanda (to appear): "Software Citation in the Digital Humanities". In: *DH 2023*.

*Collaboration as Opportunity*. Conference abstracts, Graz, 10.-14.7.2023.

Lamprecht, Anna-Lena / Garcia, Leyla / Kuzak, Mateusz / Martinez, Carlos / Arcila, Ricardo / Del Pico, Eva M. / Dominguez Del Angel, Victoria et al. (2020): "Towards FAIR principles for research software." In: *Data Science* 3 (1): 37–59.

<https://doi.org/10.3233/DS-190026>.

Smith, Arfon M. / Katz, Daniel S. / Niemeyer, Kyle E. / FORCE11 Software Citation Working Group (2016): "Software citation principles." In: *PeerJ Computer Science* 2:e86. <http://dx.doi.org/10.7717/peerj-cs.86>.

## About the authors

**Fernanda Alvares Freire** ([0000-0002-6414-5212](https://orcid.org/0000-0002-6414-5212)) is a research associate at the University of Rostock. Her work currently focuses on the development of study programs in the field of Digital Humanities. Additionally, she works on digital research methods such as social network analysis, data mining, and general data management and exploration approaches.

**Anne Fergner** ([0000-0002-1382-2658](https://orcid.org/0000-0002-1382-2658)) is a Research Associate at the University of Paderborn and University of Duisburg-Essen. Currently she mainly works on sustainable research tools and software development in the context of NFDI4Culture and FAIR research data management at the project MuMoCorp.

**Ulrike Henny-Krahmer** ([0000-0003-2852-065X](https://orcid.org/0000-0003-2852-065X)) is junior professor for Digital Humanities at the University of Rostock. She is Co-convenor of the DHd working group on Research Software Engineering (DH-RSE). Her research focuses on digital scholarly editions, digital text analysis, and evaluation and sustainability of DH research outputs.

**Daniel Jettka** ([0000-0002-2375-2227](https://orcid.org/0000-0002-2375-2227)) is research associate at the Universität Paderborn and works on sustainable software development and research infrastructure implementation in NFDI4Culture. Previously, he worked on repository management, interfaces for spoken language corpora, and digital infrastructure for the documentation of endangered languages.

# An edition of the writing *Die Wagnerfrage* in the online portal of the Joachim Raff Archive

Severin Kolb, SLUB Dresden

Poster

**Keywords:** Digital Musicology, Index of Writings, Edition of Writings, TEI

## Abstract

With his 1854 publication *Die Wagnerfrage* ("The Wagner Question") and the music drama *Samson*, the composer, music writer, and pedagogue Joachim Raff, who worked in Weimar as Franz Liszt's assistant, sought to place himself on an equal footing with his two more famous colleagues. The result is a multifaceted and erudite, but also controversially received book that bundles digressions into music history, music theory, contemporary aesthetics, and drama theory and dissects Wagner's *Lohengrin* with a sharp, scholarly eye. In parallel to the author's dissertation project on Raff's reception of Wagner in the 1850s, an index of writings with the available metadata on numerous newly discovered texts (and the possibility for the later additions of editions) as well as the historical-critical edition of the *Wagnerfrage* (with commentaries) as a prototype in TEI are to be published in the online portal of the Joachim Raff Archive. On the one hand, the digital infrastructure of the portal, which has been created on the model of the Carl Maria von Weber Complete Edition (indexes of persons, institutions, correspondence, and works), is to be reused and expanded in order to access the new content from a wide variety of angles. The rich source material and the digital format also make it possible to link the edited text in the most diverse directions (e.g., to the digital copies of the literature used by Raff, which are often already created according to IIIF standards, to the corresponding excerpts, or to Wagner's scores and piano reductions). In particular, the possibilities offered by IIIF annotations will be evaluated during the course of the project. The poster contains basic considerations on data models for the index of writings and the editions as well as a first concept for the re-use of sources available in third party repositories in order to make Raff's intellectual world visual.

## About the author

**Severin Kolb** studied musicology, religious studies and hermeneutics at the universities of Zurich, Heidelberg and Cologne. In the fall of 2018, the Joachim Raff Archive, which he co-founded, opened its doors, and he has served as its scientific director ever since. Since June 2022 he has been a research associate in the DFG project *Digital Liszt Sources and Works Catalogue* (SLUB Dresden).



# An MEI-TEI framework for the description of musical prints

Christine Siegert, Beethoven-Haus Bonn, Germany

Kristina Richts-Matthaei, Universität Paderborn, Germany

Long Paper

**Keywords:** Music encoding, Historical prints, Beethoven

## Abstract

In traditional music editions, printed sources often are underrepresented. Especially few attention is drawn on single printed copies, which, however, seems to be crucial for the aim of transparency, namely because copies that at a first glance seem to be identical, in reality are not. In our ongoing project, *Das Handwerk des Verlegers – Untersuchungen zu Entstehungsprozessen von Beethoven-Originalausgaben*, financed by the Fritz Thyssen Stiftung für Wissenschaftsförderung, we therefore focus on historical music prints. They were foremost presented in parts, and not in score format.

Our detailed study of up to now 15 authorized part editions of chamber music and symphonic works by Ludwig van Beethoven (201 copies altogether) has the following results: in 66 % (10 editions), the number of correction processes differ from part to part. To give two examples: In the Piano Trios op. 1, in the piano part we can trace three stages of the plates, whereas the violin and violoncello each have four. In the Sextet op. 81b, violin I has three stages of the plates, violin II five, the viola and horns only a single one, and the violoncello two. Moreover, in bibliographical entities, i.e., the single copy in a library, the combination of the parts is not always chronologically consistent; one part of it can be of a late stage of the plates whereas the other one can be of an early stage.

The paper will discuss the framework developed mainly by Till Reininghaus, Kristina Richts-Matthaei, and Daniel Röwenstrunk to describe these sources. On a first level, the framework combines MEI files for the printed editions themselves with TEI files for related data: biographical information on persons (composers and authors of texts, dedicatees etc.), institutions (e.g., publishing houses), places and bibliographical information. The files can easily be linked via automatically

generated IDs. Every file can be, of course, and is enriched with new editions included in the dataset. With this information, we are able to build a growing network around the prints in consideration.

A special challenge was to construct an adequate file structure to encode heterogeneous sets of parts. For the work entries, we could use the structure which the MEI header provides as a standard, itself indebted to the FRBR model. As usual, every edition is considered a `<manifestation>` (encoded in a `<manifestationList>`) where we can provide detailed information of the genesis of the print, its manufacturing process etc. Each `<manifestation>` has (or can have) different `<items>` (single copies, encoded in an `<itemList>`) with their individual characteristics as handwritten additions, stamps etc. The different stages of the plates with a detailed description of the changes, have, however, to be documented separately for the reasons mentioned above. For this purpose, we use again the `<manifestation>` element, encoded in a `<componentList>`. The single copies and their separate parts are then linked with `<relations>` of a `<relationList>`. The user has therefore access to the different copies as well as to the different stages of the plates. With this structure, we hope to contribute to a more detailed consideration of music prints in digital as well as traditional music philology.

# Bellini Digital Correspondence meets MEI

Laura Mazzaguo, Università di Pisa, Italy

Pietro Sichera, Ministero dell'Istruzione e del Merito

Daria Spampinato, Istituto di Scienze e Tecnologie della Cognizione - CNR-Catania, Italy

Angelo Mario Del Grosso, Istituto di Linguistica Computazionale "A. Zampolli" - CNR-Pisa, Italy

Long Paper

**Keywords:** Digital scholarly edition, Bellini Digital Correspondence, Vincenzo Bellini, TEI-MEI, multimedia museum

## Abstract

Bellini Digital Correspondence (BDC) is the digital scholarly edition constituted by the 40 autograph letters of *Vincenzo Bellini* (1801-1835) held at the *Museo Belliniano in Catania*, Italy. The project was conceived and developed by the National Research Council of Italy and it aims at providing a web platform (<http://bellinicorrespondence.cnr.it>) for digital philology of Bellini's texts. At the same time, it is addressed both to specialists and to occasional museum visitors, who can consult the edition by means of a touchscreen located within the museum tour. BDC is a completely open source and open access project and it is the result of a multidisciplinary initiative which implements a distributed, collaborative and cooperative editing model.

The digital edition is an image-based one and it has been implemented following the TEI guidelines. The web-based presentation of the edition has been achieved through EVT software (<http://evt.labcd.unipi.it/>) (Figure 1). Particular attention was paid to the management of metadata concerning the description of the correspondence, which were encoded through the <correspDesc> tagset within the *profileDesc* section of the *teiHeader*. The TEI *Manuscript Description* module and the <facsimile> tagset also provided an encoding mechanism for describing the primary source and linking the facsimile regions of interest with the corresponding text. The editorial process went through several phases, including one dedicated to the harmonisation and standardisation of TEI encoding, during which a number of tools – reusable in similar encoding initiatives – were developed

with the aim of ensuring the highest level of uniformity and consistency within the edition. The edition provides both diplomatic and interpretive levels based on the recent work made by Seminara (2017).

In addition to a large corpus of correspondence, the Museum safeguards numerous musical documents: these include 58 manuscripts (41 autographs) containing music by Bellini (2100 sheets ca.). These documents provide significant evidences of different genetic stages concerning all of the Catanian composer's operas, from 'Adelson and Salvini' (1824) to 'I Puritani' (1834). In addition to their location in the Museo Belliniano in Catania, letters and manuscript scores share several points (references to names of persons, entities and places, references to operas and some arias in particular, references to specific compositional processes reflected in the drafts of the manuscript scores, etc.) but, unlike the correspondence, most of the Museum's musical resources lack a scholarly digital edition.

The MEI encoding of some of these documents can be an important advancement towards the investigation of Bellini's oeuvre, especially in relation to the ongoing philological studies and formal ontologies. In particular, the encoded musical documents can be further exploited to analyse the composer's creative process, according to the approach successfully used in other similar projects (<https://beethovens-werkstatt.de/>). Moreover, thanks to the MEI encoding, the BDC project would provide an interactive and functional digital representation of different editorial stages made by Bellini to produce his text, such as, for example, the so-called 'Studi giornalieri' (Figure 2) catalogued as 'MM.B.36': sketches of melodies, textual annotations, and first draft of musical motifs partly included in the 'Puritani' and already covered in recent studies (Mantica 2020). Module 11 (*Scholarly Editing with MEI*) and Module 8 (*Lyrics and Performance Directions*) of MEI Guidelines will be particularly valuable for primary source encoding.

In our opinion, the edition of these musical texts and the digital representation of their primary sources are an interesting case study to investigate Bellini's working method, as described in various letters written by the composer. The short-term objective is to combine the edition of the letters (published in May 2023 by CNR Edizioni) with a web-visualisation proposal focused on a selection of fragments derived from this musical material, in order to formulate a working hypothesis that, by considering an integration of the museum's different artifacts, may be of interest and usefulness to Bellini scholars. The future prospects of the project will, therefore, be oriented towards a more comprehensive integration between text and

music within the edition, with a view to a potential extension of the visualisation tool to include musical content as well.



Fig. 1: BDC digital edition, first \*recto\* leaf of LL1.16 letter

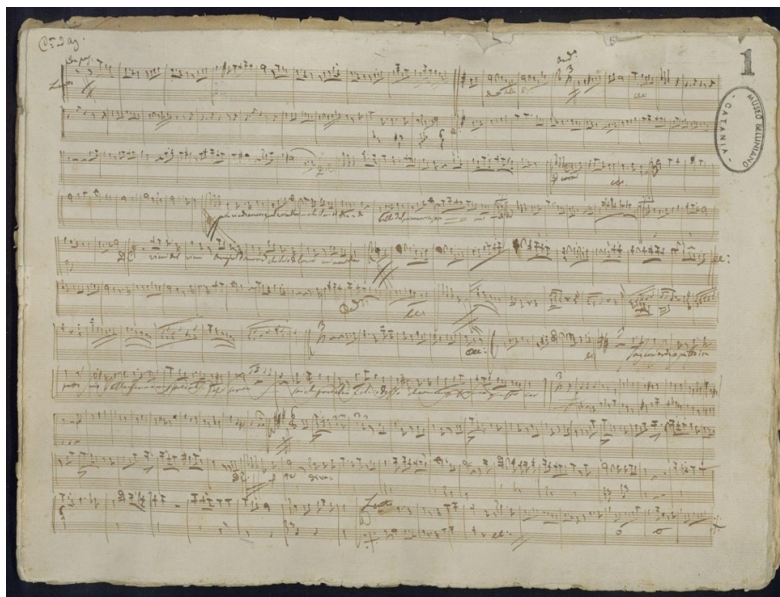


Fig. 2: 'Studi giornalieri', first leaf (MM.B.36, Museo Belliniano, Catania, Italy)

## Bibliography

Bellini, Vincenzo. 2017. *Carteggi*. Seminara, Graziella (Ed.). Firenze: Leo S. Olschki Editore.

Cristofaro, Salvatore; Spampinato, Daria. "[OntoBellini: Towards an RDA Based Ontology for Vincenzo Bellini's Cultural Heritage](#)". 2019. In A. Barton, S. Seppälä, D.

Porello (Eds.), *Proceedings of the Joint Ontology Workshops 2019*, Graz (Austria), September 23-25, volume 2518 of CEUR Workshop Proceedings.

Del Grosso, Angelo Mario; Capizzi, Erica; Cristofaro, Salvatore; De Luca, Maria Rosa; Giovannetti, Emiliano; Marchi, Simone; Seminara, Graziella; Spampinato, Daria. 2019. "[Bellini's Correspondence: a Digital Scholarly Edition for a Multimedia Museum](#)". In *Umanistica Digitale* 3(7): 23-47.

Geertinger, Alex Teich. 2021. "Digital Encoding of Music Notation with MEI". In M. S. Bue, A. Rockenberger (Eds.), *Notated Music in the Digital Sphere. Possibilities and Limitations* 15, of *Nota bene - Studies from the National Library of Norway*. 35-56. Oslo: National Library of Norway.

Mantica, Candida Billie. 2020. "Gli 'studi giornalieri' di Bellini «sviluppati con effetto» nei Puritani". In *Bollettino di Studi Belliniani* 6: 29-73.

Pellino, Santa; Sichera, Pietro; Del Grosso, Angelo Mario; Spampinato, Daria. 2022. "Dalla codifica alla fruizione: l'edizione digitale Bellini Digital Correspondence". In F. Ciraci, G. Miglietta, C. Gatto (Eds.), [AIUCD 2022 - Proceedings](#). *Quaderni di Umanistica Digitale*: 163-8.

## About the authors

**Laura Mazzaguffo** graduated in Musicology from the University of Pavia, and she is currently pursuing a Digital Humanities degree at the University of Pisa while actively contributing to the *Bellini Digital Correspondence* project. Her research interests encompass text encoding and music philology.

**Pietro Sichera** was a Research Fellow at the Istituto di Scienze e Tecnologie della Cognizione (ISTC) – CNR in Catania, Italy. He earned his degree in Computer Science from the University of Catania and obtained his PhD degree in Lexicography and Semantics of European Literary Language, furthering his expertise in the study of the Italian language through the concordance method. He has developed proprietary software for text analysis on Italian poets of the 20th century. Currently, he teaches Computer Science in Secondary Schools.

**Daria Spampinato** is a researcher at the CNR ISTC in Catania, Italy, specialising in Computer Science with a focus on Digital Humanities. She is a member of the Board of the Italian Association for Humanities and Digital Cultures. Currently, her

research interests cover Digital Philology and Digital Epigraphy, and she is actively engaged in the *Bellini Digital Correspondence* project.

**Angelo Mario Del Grosso** is a software engineer; he earned his PhD in Information Engineering with a thesis on the design of software components for textual scholarship. Since 2019, he has been a confirmed researcher at the CNR-ILC in Pisa. His main research activities encompass the representation and processing of philological resources. His interests are as much aimed at the formal modelling of text as at the computational analysis of textual documents. He is involved in different research projects both nationally and internationally and he teaches text encoding at the University of Pisa.

# Between bespoke customization and expressive interface: a reflection from the Frankenstein Variorum project

Elisa Eileen Beshero-Bondar, Penn State Erie, The Behrend College, United States of America

Raffaele Viglianti, University of Maryland

Yuying Jin, Penn State Erie, The Behrend College, United States of America

Long Paper

**Keywords:** non-hierarchical structures, overlap, xml data structure, interface, production pipeline

## Abstract

Perhaps one of the least appreciated and most definitive challenges of working with the TEI Guidelines is their requirement that scholars interpret and adapt them to the bespoke requirements of their text scholarship. It is particularly challenging to explicate overlapping hierarchies with TEI, for example:

- surface and page vs. chapter, stanza, margin notes spanning pages,
- sentence or clause vs. verse line,
- structural divisions that do not align in different versions of a text.

The TEI Guidelines' Chapter 20 about Non-Hierarchical Structures perhaps best exemplifies the challenges and possibilities of our attempts to reconcile different ideas of hierarchy in the texts that we study and curate (in our case a page-by-page encoding of a manuscript with structurally encoded print editions).<sup>[1]</sup> Our primary interest in studying these texts may be to investigate how their conflicting hierarchies were expressed at different manifestations—perhaps in alternate manuscripts or in the publication history.

The TEI Guidelines' Chapter 20 comments on the challenges to processing when we attempt to model multiple hierarchies, especially when they attempt this in inline ways. But the Guidelines largely (and purposefully) leave scholars to their own



devices for processing. We determine for ourselves, based on our available resources, what processing we can apply to express our text data modeling. Yet the processing methods we apply can either extend or limit the expressiveness of our scholarly projects in the data visualizations and digital edition interfaces we can design. The processing decisions we make will determine whether and how we can share the meaningful conflict points in multiple structures, versions, or ways of interpreting a text.

Based on our efforts with the Frankenstein Variorum project since 2017, we understand conflicting hierarchies to pose a challenge whose resolution should reflect our research questions and our scholarly document data modeling. In the context of our research questions, we will review our processing and production pipeline, including:

- Document analysis of multiple digital editions of Frankenstein to locate points of alignment and determine normalization strategies,
- XSLT required to process conflicting hierarchies by first flattening all versions and then "raising" edition files with marked variants,
- The recognition of the document data structure and its normalization in the interface processing,
- The design of a static interface that displays and foregrounds variation across conflicting hierarchies.

The pipeline processing that we devised to design the interface has changed from our first proof-of-concept launch of the website to our current effort. The decisions to work with JavaScript frameworks (React with Astro's CETELcean plugin) determine how we visualize data from our stand-off TEI "spine," which stores information about our edition files across their conflicting structures. Our pipeline is certainly "bespoke"—custom-fit to the distinct points of comparison across multiple encoded versions. The step-by-step implementation of this bespoke production pipeline is something we need to revisit, correct, update, reflect on as the basis of the important work of *interchange* that mediates across distinct data models.<sup>[2]</sup> While the interface too often represents a compromise with available tooling,<sup>[3]</sup> a production pipeline designed for interchange may open up possibilities to express the TEI in the interface.

## About the authors

Elisa Beshero-Bondar, [eeb4@psu.edu](mailto:eeb4@psu.edu), <https://orcid.org/0000-0002-1709-2314>  
(Penn State Behrend)

Raffaele Vigiante, [rviglian@umd.edu](mailto:rviglian@umd.edu), <https://orcid.org/0000-0003-3094-139X> (U. Maryland)

Yuying Jin, [yxj5181@psu.edu](mailto:yxj5181@psu.edu) (Penn State Behrend)

## Notes

---

1. Chapter 20 of the current TEI P5 Guidelines: <https://tei-c.org/release/doc/tei-p5-doc/en/html/NH.html>. ↵
2. On the significance of interchange for TEI projects, see Bauman, Syd. "Interchange vs. Interoperability." Proceedings of Balisage: The Markup Conference 2011. Balisage Series on Markup Technologies, vol. 7 (2011). <https://doi.org/10.4242/BalisageVol7.Bauman01>. Two of the authors of this proposal have previously addressed this topic at length in Beshero-Bondar, Elisa E., and Raffaele Vigiante. "Stand-off Bridges in the Frankenstein Variorum Project: Interchange and Interoperability within TEI Markup Ecosystems." *Balisage Series on Markup Technologies*, vol. 21 (2018). <https://doi.org/10.4242/BalisageVol21.Beshero-Bondar01>. ↵
3. Peter Robinson has addressed the problems of interfaces obscuring scholarly work in his presentation "Why Interfaces Do Not and Should Not Matter for Scholarly Digital Editions" *Digital Editions as Interfaces*. Graz, September 2016, <https://www.slideshare.net/PeterRobinson10/why-interfaces-do-not-and-should-not-matter-for-scholarly-digital-editions>. For a counterpoint on the importance of the interface to expressing the data model, see Andrews, Tara L. and van Zundert, Joris J. (2018). "What Are You Trying to Say? The Interface as an Integral Element of Argument," in: *Digital Scholarly Editions as Interfaces*, ed. Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber, and Gerlinde Schneider, Herstellung und Verlag, 2018, pp. 3-33. ↵

# Big (Work) Data: Making and Using MEI for Composers' Digital Work Catalogues

Jonathan Gammert, Akademie der Wissenschaften und der Literatur Mainz, Germany

Joshua Neumann, Akademie der Wissenschaften und der Literatur Mainz, Germany

Kristina Richts-Matthaei, Universität Paderborn, Germany

Poster

**Keywords:** Work Catalogues, Datasets, Metadata, Collaboration

## Abstract

Interrogating any source philologically requires first the collection or creation and curation of data, whether analogue or digital. This praxis requires increasingly interdisciplinary methodologies as musicology continues its digital turn—methodologies that are most quickly applied to a project through collaboration. A natural output of such projects undertaken at the scale of collected works is a catalogue documenting metadata, which account for an author's or composer's output. These catalogues can also establish intellectual histories of these corpora and, for major figures, (i.e. – those with voluminous works and expressions or manifestations thereof), they establish paradigms of what constitutes "Big Data" in these fields.

Moreover, they invite examination of the terminology used. Is it simply "Data" and "Metadata" that researchers can enrich through authority data? Does it expand to some other qualifier such as "Big Data," "Deep Data," or "Broad Data," even with the various challenges inherent to each? However one qualifies these datasets can understandably affect the ways in which creators, curators, and users interact with them. Considerations in this regard, combined with assessment of the nature of the dataset in question, whether extant or projected, and design thinking practices of end-user/stakeholder input lead to individually tailored solutions.

Two collaboratively born datasets are currently in progress in the context of NFDI4Culture, the national Consortium for Research Data on Material and Immaterial Cultural Heritage in Germany. Each one focuses on (research) data management, metadata, documentation, and digital transformation. One

collaboration in early stages is creating conceptual and digital infrastructures for the transformation of metadata for the works of Georg Philipp Telemann. A unique challenge in this interdisciplinary cooperation is how to handle, at a fundamental level, a myriad of idiosyncrasies unique 18th-century compositional praxis as they emerge in Telemann's c. 5,000 works. This project establishes the digital infrastructure through the Metadata editor and repository for MEI data (MerMEId) for the online Telemann catalogue, which will soon be published in phases by the Telemann-Zentrum Magdeburg. Another dataset, nearing publication in Radar4Culture, contains approximately 33,000 files (c. 3,000 work files and c. 30,000 source files) resulting from the *Joseph Haydn Werke*. Collaboration was essential in this project to preserve data contents while transforming from a single-use schema into MEI.

These datasets, primarily comprised of metadata, constitute two contributions of value for musicologists and the MEI and TEI communities. The poster presents a collaborative approach to data transformation into MEI from previous formats, addresses specific challenges associated therewith, and highlights a cooperation model that actively contributes to broader digital philology communities of text and music.

# Brainstorming on the Encoding of Eastern Neumes: Middle-Byzantine Notation as a Case-Study

Elsa De Luca, Nova University Lisbon, Portugal

Maria Alexandru, Aristotle University of Thessaloniki

Ichiro Fujinaga, Schulich School of Music, McGill University, Canada

Long Paper

**Keywords:** Byzantine neumes, Orthodox chant, Music Encoding, MEI Neumes Module, Tablature

## Abstract

The effort to encode neumatic notations has primarily focused on Western European repertoires, with little attention given to the neumes employed in the East. To address this gap, we propose to apply our knowledge of the MEI Neumes Module and Eastern notations to raise awareness on some methodological issues that have hindered modern scholars from encoding Eastern neumes. Our case study is the Middle Byzantine notation, employed from the mid-12th-century to about 1815 across a large geographical area encompassing mainly Eastern Europe and part of the Middle East. This notation has been chosen since it has many points of contacts with Western notations but, at the same time, its basics differ radically from what we encounter in the West.

In the East, music scripts are of the 'articulatory' type, serving as a guide on how to make music, almost similar to a tablature style of notation. In Western notation the performance is captured mostly within the neume. In the East notation is context sensitive and its interpretation depends on different factors, such as the genre and style of *melopoeia*, the mode, the melodic formula, the place within the musical phrase, the poetic text, and the liturgical frame. This fundamental difference has also an impact on the kind and number of signs employed in the notations. Middle-Byzantine notation offers an impressive number of combinations and positioning of signs including neumes that provide specific interval information and subsidiary or cheironomic signs indicating the musical 'Gestaltung', which implies rhythmical properties, grouping, and phrasing of the interval signs, dynamics, ethos of the

formulas, as well as different levels of ornamentation and melodic expansion, according to the context. Possibly, the most difficult feature to be captured in the encoding of Middle-Byzantine notation is the relative position of the signs in the notational space; indeed, their relative positioning (above, below, next) can change the overall interpretation of all the related signs.

When considering the graphical appearance and semantic of neumes, we should adopt both approaches to the encoding of Middle-Byzantine notation. It is crucial to be able to capture in the encoding even the finest details because this notation is very precise and conveys musical meaning through the tiniest graphical feature. On the other hand, encoding the semantic of the neumes should remain possible and open since, besides the basic intervallic structure of the pieces, the other dimensions of the music have to be interpreted in the light of oral tradition. Therefore, it would be desirable to include information on historically informed practice, which is available through musical treatises, manuscripts, and the oral tradition of Byzantine Chant. Middle-Byzantine neumes are therefore rich in metadata and their encoding should also include the names of the neumes in Greek, possibly with transliteration, translation, and etymology.

We hope that our presentation can actively contribute to the ongoing discussions on encoding of non-Western early notations, particularly the Middle-Byzantine notation, and encourage a collaborative effort to find the most effective way to encode Eastern neumes.

## Bibliography

De Luca, Elsa, and Haig Utidjian. 2023. Challenging the MEI Neumes Module: Encoding Armenian Neumes. In *Proceedings of the Music Encoding Conference*, Dalhousie University, 19–22 May 2022, ed. by Jennifer Bain and David M. Weigl. Forthcoming.

De Luca, Elsa, Jennifer Bain, Inga Behrendt, Ichiro Fujinaga, Kate Helsen, Alessandra Ignesti, Debra Lacoste, and Sarah Ann Long. 2019. Capturing Early Notations in MEI: The Case of Old Hispanic Neumes. *MusikTheorie-Zeitschrift für Musikwissenschaft* 3: 229–49.

De Luca, Elsa, Jennifer Bain, Inga Behrendt, Ichiro Fujinaga, Kate Helsen, Alessandra Ignesti, Debra Lacoste, and Sarah Ann Long. 2019. Cantus Ultimus' MEI Neume Module and its Interoperability Across Chant Notations. Abstract. *Music Encoding Conference*, Vienna. Available at <https://music->

[encoding.org/conference/abstracts/abstracts\\_mec2019/MEI%202019\\_De%20Luca%20et%20alii.pdf](https://encoding.org/conference/abstracts/abstracts_mec2019/MEI%202019_De%20Luca%20et%20alii.pdf).

Alexandru, Maria. 2017. *Paleography of Byzantine Music: Scientific and artistic quests (in Greek: Παλαιογραφία Βυζαντινής Μουσικής. Επιστημονικές και καλλιτεχνικές αναζητήσεις)*. Athens: Hellenic Academic Ebooks.  
<https://repository.kallipos.gr/handle/11419/6487>

Siklaidis, Nikolaos, and Maria Alexandru. 2022. Font Design of Psaltic (Byzantine) Notation for Greek Musical Repertories. In *Proceedings of the Music Encoding Conference, Alicante 2021*. (poster και αντίστοιχο κείμενο):  
<https://hcommons.org/deposits/item/hc:46011/> (3.12.2022).

Troelsgård, Christian. 2011. *Byzantine Neumes. A New Introduction to the Middle Byzantine Musical Notation*. MMB, Subsidia IX. Copenhagen: Tusculanum Press.

## About the authors

**Elsa De Luca** is an early music scholar pursuing research on medieval notations at the Centre for the Study of the Sociology and Aesthetics of Music, NOVA University Lisbon. She is the PI of the interdisciplinary project *Echoes from the Past: Unveiling a Lost Soundscape with Digital Analysis*.

**Maria Alexandru** is a scholar of Byzantine Music and head of the School of Music Studies of the Aristotle University of Thessaloniki, as well as chair of the IMS Study Group *Music of the Christian East and Orient*.

**Ichiro Fujinaga** is Canada Research Chair in Music Information Retrieval and Professor in the Music Technology Area of the Department of Music Research at the Schulich School of Music, McGill University. He has been working on encoding neume notation into MEI since 2010.

# Building sustainable infrastructure for scholarly communities

Peter Stadler, University of Paderborn, Germany

Anne Ferger, University of Paderborn, Germany

Long Paper

**Keywords:** TEI format, MEI format, infrastructure, sustainability, software, API, REST, transformations

## Abstract

While the long term preservation and availability of research data has already seen some discussion and solutions, the sustainability of research software and research infrastructures is still less discussed. It mostly depends on individual institutions and people that bring together the technical, social, and economic (i.e. funding) requirements. The paper discusses the development of the TEIGarage/MEIGarage and its service availability as a model for sustainable infrastructure in scholarly communities, especially highlighting the need of building communities for and collaborating on maintaining research software infrastructures and the advantages of virtualisation and central deployment.

The popular term of *sustainability* often means the environmental factor, but there is also economic and social sustainability (see Sverdrup and Svensson 2005 and Venters et al. 2014). Ecological sustainability for research software and infrastructure (e.g. green computing) is recently discussed in the context of Digital Humanities (see Baillot et al.), and while we will not focus on this aspect here, we consider it a crucial aspect. In the context of research software and infrastructure technical sustainability is the most widely discussed factor, but economic/financial and organizational sustainability and funding also play a crucial role. The practical question we want to discuss in this abstract is how the “Garage” infrastructure can be sustained in the best way.

TEIGarage<sup>[1]</sup> and MEIGarage<sup>[2]</sup> (formerly OxGarage) is a long-living software project used in various contexts (see e.g. Stadler et al. 2022). The software offers a framework for community specific data processing.



Allowing for collaboration on the framework as well as on its maintenance and deployment as a service are crucial for tackling the workload in small research communities such as the TEI and MEI communities. With their open and general framework the Garages offer infrastructure for scholarly communities.<sup>[3]</sup> While the Text Encoding Initiative constituted itself as an infrastructure (see Burnard 2013), the Garage framework represents a technical form of infrastructure.

A central deployment of research software offers multiple benefits as opposed to local installations, especially in the light of sustainable and reproducible research. A central low-threshold interface enables scholars without technical expertise to exploit the functionality, e.g. students. By offering a reliable API endpoint further tools can easily build on the services (see e.g. Roma<sup>[4]</sup>). One central deployment of the software also ensures the reproducibility of the e. g. conversion which increases the traceability of the research. In the case of TEIGarage and MEIGarage this also means the utilized Guidelines and Stylesheets can be retraced.

But these benefits don't come without costs. Resources such as hardware and basic software infrastructure are needed for container orchestration and virtualisation, as well as maintainers for the infrastructure. This also means work on the software apart from the code base, but on its virtualisation and containerisation.

To be able to cover these costs as small research communities, forces need to be joined. In the example of the TEIGarage and MEIGarage not only the underlying framework can be reworked collaboratively, also the centralized deployment and virtualisation can be combined. A facilitating factor is the technical sustainability of the code the software is based on. Sustainable code and software does ideally need less resources for maintenance. Measures have been taken for the Garage framework to become more sustainable and FAIR in the sense of Hasselbring et al. (2020), Anzt et al. (2021) and Hong et al. (2022).

In conclusion and from our practical experience, costs (working hours as well as software and hardware costs) need to be shared between small research communities for software to be sustained over a long period of time.

## Bibliography

Anzt, H. *et al.* (2021) 'An environment for sustainable research software in Germany and beyond: current state, open challenges, and call for action'. F1000Research. Available at: <https://doi.org/10.12688/f1000research.23224.2>.

- Baillet, Anne *et al.*: „Digital Humanities and the Climate Crisis. A Manifesto“, *Foregrounding the Climate Crisis Within Digital Humanities Work*, <https://dhc-barnard.github.io/dhclimate/> (accessed 24.04.2023).
- Burnard, L. (2013) 'The Evolution of the Text Encoding Initiative: From Research Project to Research Infrastructure', *Journal of the Text Encoding Initiative* [Preprint], (Issue 5). Available at: <https://doi.org/10.4000/jtei.811>.
- Hasselbring, W. *et al.* (2020) 'From FAIR research data toward FAIR and open research software', *it - Information Technology*, 62(1), pp. 39–47. Available at: <https://doi.org/doi:10.1515/itit-2019-0040>.
- Hong, N.P.C. *et al.* (2022) *FAIR Principles for Research Software (FAIR4RS Principles)*. Available at: <https://doi.org/10.15497/RDA00065>.
- Penzenstadler, B. (2013) 'Towards a definition of sustainability in and for software engineering', in *Proceedings of the 28th Annual ACM Symposium on Applied Computing*. New York, NY, USA: Association for Computing Machinery (SAC '13), pp. 1183–1185. Available at: <https://doi.org/10.1145/2480362.2480585>.
- Stadler, Peter, Feger, Anne and Röwenstrunk, Daniel (2022) 'From OxGarage to TEIGarage and MEIGarage'. Available at: <https://doi.org/10.5281/ZENODO.7061525>.
- Sverdrup, H. and Svensson, M.G.E. (2005) 'Defining the Concept of Sustainability - a Matter of Systems Thinking and Applied Systems Analysis', in M.-O. Olsson and G. Sjöstedt (eds) *Systems Approaches and Their Application: Examples from Sweden*. Dordrecht: Springer Netherlands, pp. 143–164. Available at: [https://doi.org/10.1007/1-4020-2370-7\\_6](https://doi.org/10.1007/1-4020-2370-7_6).
- Venters, C. C. *et al.* (2014) 'Software Sustainability: The Modern Tower of Babel', *RE4SuSy: Third International Workshop on Requirements Engineering for Sustainable Systems*. Edited by B. Penzenstadler, M. Mahaux, and C. Salinesi. CEUR. Available at: <https://eprints.whiterose.ac.uk/84941/>.

## About the authors

**Anne Feger** [0000-0002-1382-2658](https://orcid.org/0000-0002-1382-2658) is a Research Associate at the University of Paderborn and University of Duisburg-Essen. Currently she mainly works on sustainable research tools and software development in the context of NFDI4Culture and FAIR research data management at the project MuMoCorp.

**Peter Stadler** [0000-0002-2544-1481](https://orcid.org/0000-0002-2544-1481) is a Research Associate at the University of Paderborn. He currently works on the digital edition of Carl Maria von Weber and is a long term enthusiast of both MEI and TEI. He has been a member of the TEI Council for several years.

## Notes

---

1. <https://teigarage.tei-c.org/> ↩
2. <https://meigarage.edirom.de/> ↩
3. While this currently means the MEI and TEI communities, there will hopefully be further Garages by further research communities. ↩
4. Both “Romaantiqua” (<https://romaantiqua.tei-c.org/>) and the new “Roma” at <https://roma.tei-c.org/> (also known as “RomaJS” or “Romabeta”) use TEIGarage (formerly OxGarage) as backend. ↩

# Building Workflows for HTR to TEI Up-Conversion and Enhancement

James Cummings, Newcastle University

Diane Jakacki, Bucknell University, United States of America

Alexandra Healey, Newcastle University, United Kingdom

Carrie Pirmann, Bucknell University, USA

Ian Johnson, Newcastle University, United Kingdom

Evie Jeffrey, Newcastle University, United Kingdom

Valentina Flex, Newcastle University, United Kingdom

Short Paper

**Keywords:** TEI XML, Handwritten Text Recognition, Workflows, HTR

## Abstract

At the TEI2022 conference, we introduced the Evolving Hands project which was undertaking three case studies ranging across document forms to demonstrate how HTR-based workflows to develop TEI resources can be iteratively incorporated into curation. The case studies cover a wide range of textual forms, from handwritten work from the 19th-20th century UNESCO-recognised *Gertrude Bell Archive* of letters and diaries, or 20th century artistically-printed feminist zines, to highly structured and scholarly edited legacy printed material of the *Records of Early English Drama* (REED) project. The printed material used as samples are all in categories where using traditional OCR would result in a substantial loss of intellectual content. By covering a wide variety of periods and document forms the project wants to explore how best cultural institutions might use HTR to convert their collections into TEI files and then further approaches for up-conversion and enhancement of these into data-rich TEI editions.

Across the varying case studies different approaches have been used, such as those who attempted to do as much tagging as possible in Transkribus' online app, those who used LEAF-Writer to add enhancements liked named entity LOD, or those who processed the output with XSLT to preserve or enhance structured information before editing it further.

This short paper will present some of the interim results of developing these workflows for converting HTR-processed text into TEI P5 XML and its further enhancement. Our goal in doing so is to pass on some of the insights and lessons learned during the development phase of the project, as well as to provide a starting point for further discussions and future collaborations in the field of HTR-based workflows for TEI curation. Ultimately, the project hopes to contribute to the ongoing development of best practices for using HTR for the creation of TEI editions.

## About the authors

**James Cummings** is the Reader for Digital Textual Studies and Late Medieval Literature for the School of English Literature, Language, and Linguistics at Newcastle University. He is the Newcastle PI for the Evolving Hands Project and on the TEI Board of Directors.

**Diane Jakacki** is Digital Scholarship Coordinator and Associated Faculty in Comparative & Digital Humanities at Bucknell University. She is PI of the Mellon-funded LAB Cooperative, Bucknell PI of the Evolving Hands project, and the LEAF-VRE project. She is chair of the TEI-C and Chair-Elect of ADHO.

**Ian Johnson** is Head of Special Collections & Archives at Newcastle University Library. This includes interdisciplinary digital scholarship and co-curation of our UNESCO Gertrude Bell Archive. He is co-I for the Evolving Hands project.

**Carrie Pirmann** is the social sciences librarian at Bucknell University who is optimising existing Transkribus models for the Bucknell Case Study and working with the digital libraries community.

**Alexandra Healey** is a Project Archivist in Newcastle University Special Collections & Archives. She is coordinating the use of HTR and TEI within the Newcastle team as part of the Evolving Hands project.

**Valentina Flex** is the Stillman Project Archivist working on Gertrude Bell Archive: Bell and the Kingdom of Iraq at 100.

**Evie Jeffrey** is the postgraduate assistant for the project and started working on the project as a Robinson Bequest Bursary-holder.

# Comparison of MusicXML export capabilities of different scorers

Klaus Rettinghaus, Enote GmbH, Germany

Gerrit Bogdahn, Enote GmbH, Germany

Kaspar Querfurth, Enote GmbH, Germany

Poster

Keywords: MusicXML

## Abstract

In digital musicology, the widely adopted de facto standard for creating digital critical editions is the encoding format defined by the Music Encoding Initiative. However, the usual workflow to arrive at the desired encoding involves the use of well-established WYSIWYG music notation software.

Scores are mostly prepared in the conventional way and then exported to MusicXML format to be converted to MEI afterwards. While all programs handle the export of basic features like notes, measures, clefs and key signatures quite well, there are decided differences when it comes to more complex notational features like ornaments.

We have thoroughly investigated which of the most popular programs in the latest versions offers the best export under the condition of the least manual XML post-processing effort.

We think that for upcoming edition projects this could have an impact on the workflow that might be planned.

# Computing and displaying similarity in Beethoven's vocal sketches and score for *Leonore* (1805)

Lena Katharina Alfter, University of Cambridge, United Kingdom

Short Paper

**Keywords:** computational analysis, encoding music, sketch-score-relationship, embedding similarity plots in notation

## Abstract

This project features a computational approach for analysing and rendering musical similarity between sketch and score of Beethoven's 'Gut, Söhnchen, gut' trio from the first act of the 1805 *Leonore* version of his opera. Initially, both the relevant music from the sketch and the score are encoded in MEI. Using Python, information for each musical event—be it a note, a tied note, or a rest—from these XML files is extracted before further data is added to the retrieved characteristics thereby enabling these events to be mapped onto a time-axis. Subsequently, an algorithm aligns the music from the sketch and score as well as possible with the lyrics serving as checkpoints. The result of this step is a collection of coordinate pairs containing corresponding positions from both data sets. The Python program consequently calculates a similarity value for each of these coordinate pairs with regard to their pitch, octave, duration and metrical position within the bar as a Hamming distance, yielding a value between 0 and 4. This data is then plotted in the form of a heatmap-histogram-combination with the width of each data point reflecting the duration of the corresponding musical event. Finally, making use of LilyPond's unique proportional duration notation feature, these plots are then directly embedded into a LilyPond file containing the encoded sketch transcription. This visualisation allows for an overlaid evaluation and interpretation of the relationship between the two different stages in composition. At the same time, it offers an intuitive and unbiased wide-angle perspective on the same aspect and thereby enhances not only the accessibility of a certain component of a particular musical analysis but also of the underlying primary source itself. The modular structure of

this approach provides a framework with a high degree of flexibility that could be expanded upon in future applications.



# Cracking the Code: Overcoming the Challenges of Encoding Correspondence

Dimitra Grigoriou, Austrian Center for Digital Humanities and Cultural Heritage, Austria

Short Paper

**Keywords:** correspondence, encoding, TEI, correspond-actions, representation

## Abstract

W. H. Auden (1907-1973), a prominent figure in the English-language literary landscape of the twentieth century. Despite extensive scholarly exploration of Auden's English and American periods, his life and artistic contributions in Austria remain comparatively under-explored. It was not until the early 2000s that this aspect of the poet's life began to garner scholarly attention, as evidenced by the works of Mendelson (2004) and Smith (2004). A new project at the Austrian Centre for Digital Humanities provides an open-access scholarly digital edition (SDE) that renders accessible the hitherto unattainable "working correspondence" between Auden and Stella Musulin, a writer of Welsh-Austrian origin, dating from 1959-1973 exchanged.

The decision to employ TEI for encoding the materials was predicated on its provision of conventions that facilitate the description of a text's physical and semantic structure (Burnard, 2014; Pierazzo, 2015b). Within the TEI language, the `<correspDesc>` tag affords comprehensive guidelines for encoding correspondence by specifying various types of correspondence actions (i.e., sent, received, forwarded, redirected). We aimed at making these communicative activities explicit in formal representations of the information extracted from the materials; nonetheless, in some of the documents making clear distinctions between the categories proved challenging. While our objective was to classify our data within these four categories, we released the need to supplement the existing list of `<correspAction>` types with a new type, "composed". Our analysis has revealed that the location and date of letter composition may differ from the location and date of its dispatch, which cannot be accommodated by the current encoding standards. Bibliographic resources (such as Stadler, Illetschko and Seifert 2016,12) acknowledge this peculiarity but fail, at least to our knowledge, to highlight

and include this necessary information in the metadata. The solution we propose not only addresses the specific requirements of our project but also satisfies the broader academic imperative of encoding correspondence, particularly in relation to letters from earlier historical periods which may include additional postal specificities such as postal service, stamps, etc.

## Bibliography

Burnard, Lou. 2014. *What is the Text Encoding Initiative?* OpenEdition Press.  
<https://doi.org/10.4000/books.oep.426>

Mendelson, Edward. 2005. 'The European Auden'. In: Smith, S. (ed), *The Cambridge Companion to W. H. Auden*. Cambridge: Cambridge University Press, pp. 55-67, <https://doi.org/10.1017/CCOL0521829623.005>.

Pierazzo, Elena. 2015. 'Textual Scholarship and Text Encoding'. In: *A New Companion to Digital Humanities* (pp. 307-321). John Wiley & Sons, Ltd.  
<https://doi.org/10.1002/9781118680605.ch21>

Smith, Stan. 2005. 'Introduction'. In: Smith, S. (ed), *The Cambridge Companion to W. H. Auden*. Cambridge: Cambridge University Press, pp. 1-14,  
<https://doi.org/10.1017/CCOL0521829623.001>.

Stadler Peter, Illtetschko, Marcel and Seifert, Sabine. 2016, 'Towards a Model for Encoding Correspondence in the TEI: Developing and Implementing <correspDesc>'. In: *Journal of the Text Encoding Initiative* [Online], Issue 9 | September 2016 - December 2017, Online since 24 September 2016, connection on 19 April 2019. URL: <http://journals.openedition.org/jtei/1433>; DOI: [10.4000/jtei.1433](https://doi.org/10.4000/jtei.1433).

# Creating Digital Editions with EditionCrafter

Nicholas Laiacona, Performant Software Solutions, United States of America

Short Paper

**Keywords:** digital editions open access publishing

## Abstract

EditionCrafter is a new open source tool for the creation of digital critical editions encoded in TEI XML. It can generate sophisticated diplomatic renderings of text side-by-side with a deep zoom window of the original folios. The editor can provide multiple layers of text, such as a transcription and a translation. There are also built in functions for display of editorial notes and a glossary. In this presentation, we will demonstrate EditionCrafter and how to use it.

EditionCrafter has three parts: a command line tool, a Javascript library, and a Hugo CMS theme. The command line tool converts a TEI Document into a IIIF Manifest and a collection of XML and HTML partials. The IIIF Manifest can then be provided to the EditionCrafter Javascript library to display the edition. The example Hugo theme makes it simple, you just put the files in the right place and link to them in your page's frontmatter.

TEI Documents used by EditionCrafter obey a few simple rules. First, they contain a facsimile element that links to the folio images. Each surface has an XML ID which is then referenced in one more more text or sourceDoc elements via facsimile attributes. EditionCrafter uses this information to construct the manifest and the text partials for each page.

The EditionCrafter Javascript library is a React component and can be easily added to any React based website. Additionally, we provide a version that can be included in any website with a script tag.

EditionCrafter is a generalization of code from the digital critical edition *Secrets of Craft and Nature in Renaissance France* published by the *Making and Knowing* Project at the Center for Science and Society at Columbia University.

EditionCrafter development is funded by a grant from the National Science Foundation and led by Prof. Pamela Smith.

## About the author

**Nick Laiacona** is the President of [Performant Software Solutions LLC](#). Since 2006, Performant has cultivated a specialty in critical digital editions and has served clients throughout North America and Europe. Laiacona has acted as technical lead on digital projects funded by the National Endowment for the Humanities, the Andrew W. Mellon Foundation, and the National Institutes of Health.

# Digital critical edition of Čiurlionis' piano music

**Darius Kučinskas**, Kaunas University of Technology, Lithuania; Lithuanian Academy of Music and Theatre, Lithuania

**Jūratė Janutėnaitė-Bogdanienė**, Vilnius Gediminas Technical University, Lithuania; Lithuanian Academy of Music and Theatre, Lithuania

**Short Paper**

**Keywords:** Čiurlionis, music, digital edition, critical text, database

## Abstract

Mikalojus Konstantinas Čiurlionis (1875–1911) is a pivotal figure in modern Lithuanian music. The process of editing and publishing his music has spanned over a hundred years, resulting in as many as five different cataloging systems and up to ten different editorial versions of the same piece. However, a critical text and a canonical version of the works that fulfill the needs of the modern music community are not yet available at both national and international levels.

In 2022, a project funded by the Lithuanian Research Council began to create a digital critical edition of Čiurlionis's piano pieces. This includes approximately 200 piano pieces from 1894–1910, composed half in a romantic style and half using a proto-serial compositional technique. The main aim of this project is to establish a canonical text for each piano piece and develop a tool for further research—an integrated database with the ability to search and analyze various musical aspects of the scores, such as rhythmic, melodic, harmonic, and structural features.

The canonical text will have a direct comparison with all manuscripts at the bar level and all previous editions at the page level. For this reason, all scanned manuscripts have been divided into bars (about 14,000 bars) and linked with scores online.

An algorithm for analyzing the main musical aspects will be created. The primary purpose of this tool is to enable searches and analyses of the score within one separate piece as well as across the entire database (about 350,000 notes).

The digital critical text of piano scores will be accompanied by detailed annotations and comments, as well as links to other sources and literature.

The database will be presented as an integrated online catalog, providing the ability to search for music based on primary bibliographical aspects and conduct analytical research according to main musical criteria.

## Bibliography

Berndt, Axel. *Music Performance Markup: Format and Software Tools Report*. Music Encoding Conference 2021 (<https://hcommons.org/deposits/item/hc:45983/>)

Besasparytė, Edita; Kučinskas, Darius. *Statistical analysis of musical rhythmic. Compositions for a string quartet by Lithuanian composers*. In: *Kūrybos erdvės*. 2013, Nr. 18, p. 8-18.

Besasparytė, Edita; Kučinskas, Darius. *Statistical analysis of musical rhythmic. Čiurlionis's symphonic music*. In: *Kūrybos erdvės*. 2011, Nr. 15, p. 8-17.

Desmond, Karen; Pugin, Laurent; Regimbal, Juliette; Rizo, David; Sapp, Craig; Thomae, Martha E. *Encoding Polyphony from Medieval Manuscripts Notated in Mensural Notation*. Music Encoding Conference 2021, 2021 (<https://doi.org/10.17613/xf2j-x697>)

Janeliauskas, Rimantas. *Unrecognized Musical Cycles of Mikalojus Konstantinas Čiurlionis*. Vilnius: Lietuvos kompozitorių sąjunga, 2010.

Kučinskas, Darius. *Complete Chronological Catalogue of Čiurlionis Music*. Kaunas: Technologija, 2007.

Nunokawa, Yumiko. *Contextuality of the artistic language of M. K. Čiurlionis: links and influences of cultures of the East and the West discovered in Čiurlionis' art*. Disertacija. Kaunas: Kauno technologijos universitetas, 2017.

Ros-Fábregas, Emilio. *Encoded Spanish Music Heritage through Verovio: The Online Platforms Fondo de Música Tradicional IMF-CSIC and Books of Hispanic Polyphony IMF-CSIC*. Music Encoding Conference 2021 (<https://doi.org/10.17613/m2ny-6b18>)

Short, H, Bradley, J, Vetch, P, Spence, P & Pierazzo, E, *Online Chopin Variorum Edition (OCVE)*, 2006, Data set/Database, University of Cambridge. (<http://www.ocve.org.uk>)

Vetch, P, *Chopin's First Editions Online (CFEO)*, 2007, Data set/Database, University of Cambridge. (<http://www.cfeo.org.uk/>)

## About the authors

**Darius Kučinskas** (b. 1966) is an associate professor at Kaunas University of Technology (Lithuania). He graduated from the Lithuanian Music Academy as a pianist and defended his dissertation on Čiurlionis' music in 2002. Over the last decade, Kučinskas has published *A Complete Chronological Catalogue of Čiurlionis's Music* (2007) and several urtext editions. As an active musicologist, he presents his research in Europe (Athens, Helsinki, Krakow, Paris), East Asian countries (Tokyo, Nanjing), and the United States (Stanford, Chicago, New York). Kučinskas has received BAFF (2018) and Fulbright (2022) grants for research in the US, as well as several grants from the Lithuanian Research Council and Lithuanian Cultural Council for research on national music. Since 2010, Kučinskas has been investigating ethnic music for player pianos. His research has resulted in *The Complete Catalogue of Lithuanian Piano Rolls* (2014) and the edited monograph *Ethnic Piano Rolls in the United States* (2021). In parallel, Kučinskas is conducting research on the musical culture of the Lithuanian diaspora. He regularly works in archives abroad and edits unpublished scores of Lithuanian composers.

**Jūratė Janutėnaitė-Bogdaniėnė** (b. 1988) is an associate professor at Vilnius Gediminas Technical University (Lithuania). She received a bachelor's degree in statistics in 2011, a Master of Informatics in 2014, and a PhD in mechanical engineering in 2019. She has work experience in IT system development in both the private sector and academia. Together with Darius Kučinskas, she prepared the first edition of the online Čiurlionis musical catalog. Her research areas include mathematical modeling, statistical data analysis, machine learning, and big data.

# Digital Representation of 'A Match of Crickets in Ten Rounds of Verse and Image': Text Encoding and Viewer Implementation for Japanese Poetry Match

Shizuka Fujiwara, Kyoto Women's University , Japan; Kyoto Sangyo University, Japan

Kiyonori Nagasaki, International Institute for Digital Humanities

Hiroyuki Ikuura, National Institute of Japanese Literature, Japan

Teiko Morita, Kyoto Sangyo University, Japan

Yoichi Ikura, Osaka University, Japan

Ok Matumoto, Kansai University, Japan

Kato Yumie, Nagoya City University, Japan

Poster

**Keywords:** Japanese classical poetry, Japanese Poetry Match, Japanese literature

## Abstract

Waka poem, a type of Japanese poetry, is sometimes published simply as a collection of poems, but they often appear in various contexts. The Mushi-awase, which is the subject of this presentation, was an event held in the 18th century as part of the revival of the medieval vogue, in which dioramas were created and brought together with the theme of insects that made sounds and chirped, and waka poems were added to the dioramas to compete for the contest. It was a typical example of the restoration of dynastic culture that was popular then. A beautiful picture scroll of this Mushi-awase has been recorded and preserved today, along with its diorama. The authors decided to encode this picture scroll in accordance with TEI as part of their research on this fashion.

The diorama images, wakas, and their respective ratings and wins/losses were encoded to have a reference related to each other with ID-refs. In addition, `<persName>` elements are embedded in each poet appearing in the scroll, and IDs



are assigned in anticipation of the future construction of a waka poet's name authority.

In addition, we developed a JavaScript viewer for vertical writing and the image with OpenSeadragon and jQuery.

In this structure, each `<1>` has a different poet, currently represented by `@resp`. In Japanese waka poetry, in general, it is often necessary to have a poet or author for each waka, and such a structure is generally necessary. We are currently considering if it is appropriate to use `@resp` here or if a new dedicated attribute is needed.

In this representation, we also plan to create and provide an English translation. We hope that the world will be able to enjoy this expression even more widely.

# Digitization of Ryukyuan Music: MEI of Kunkunshi Notation for Classical and Modern Songs with Sanshin

So Miyagawa, National Institute for Japanese Language and Linguistics, Japan  
Shintaro Seki, The University of Tokyo, Japan

Long Paper

**Keywords:** MEI, Ryukyuan music, kunkunshi, sanshin, tablature

## Abstract

Ryukyuan music originated in the Ryukyu Archipelago (southwest of Japan) and is an important aspect of East Asian music history (Thompson 2008). Ryukyuan music, from classical Ryukyuan music to contemporary popular music, uses a unique tablature called *kunkunshi*. It is used as the musical notation for the *sanshin*, the three-stringed instrument central to Ryukyuan music. Despite its cultural importance, *kunkunshi* sheet music faces challenges in digital preservation and accessibility. This paper proposes to use the Music Encoding Initiative (MEI) to digitize and disseminate *kunkunshi* notation to ensure its survival and continued appreciation.

MEI is an XML-based encoding standard that provides a robust framework for complex musical notation and enables accurate encoding of *kunkunshi* notation. This study outlines the adaptation of the MEI schema to the unique features of a *kunkunshi* score, including vertical writing structure, seven major characters representing tones, rhythmic values, and articulations. Also considered is the representation of microtones unique to *sanshin* music.

The *sanshin* is a three-stringed plectrum instrument and a representative instrument of Ryukyuan music, characterized by its snakeskin-covered body and distinctive tone. The *kunkunshi* notation, which effectively expresses the nuances of *sanshin* playing, is an essential element of the Ryukyuan music tradition.

To demonstrate the potential of MEI in Ryukyuan music, we are digitizing a collection of *kunkunshi* scores, based on MEI for lute tablature (Lewis 2014).

Transcription challenges such as interpreting handwritten manuscripts, deciphering obscure symbols, and reconciling different readings will be discussed.

Digitization using MEI has the advantages of digital preservation, facilitating collaboration and data sharing, and creating new opportunities for analysis and research. This paper demonstrates the potential of MEI as a powerful tool for the preservation and dissemination of didactic scores in Ryukyuan classical and contemporary popular music, facilitating the continued appreciation and study of this important cultural heritage.

## Bibliography

Lewis, Richard (2014) "Customising MEI for lute tablature," Transforming Musicology: An AHRC Digital Transformations project, the University of Oxford, <https://tm.web.ox.ac.uk/blog/customising-mei-lute-tablature> (accessed April 29, 2023).

Thompson, Robin (2008) "The music of Ryukyu." In: Alison McQueen Tokita and David W. Hughes (eds.), *The Ashgate Research Companion to Japanese Music*, pp. 303–322. London: Routledge.

## About the authors

**So Miyagawa**, a tenure-track assistant professor at the National Institute of Japanese Language and Linguistics, engages in a project on the digital preservation and revitalization of endangered languages, such as Ryukyuan and Ainu, while interested in Ryukyuan music and playing the *sanshin*.

**Shintaro Seki** is a research fellow (DC2) of the Japan Society for the Promotion of Science and a doctoral student in the Graduate School of Humanities and Sociology, The University of Tokyo. His PhD research aims to promote the study of Japanese music through an informatics-based approach.

# Digitizing Buddhist Genealogy: Encoding the Shinran Shonin Montei Kyōmyo-chō

Yoshihiro Sato, The University of Tokyo, Japan

Kiyonori Nagasaki, International Institute for Digital Humanities

Masahiro Shimoda, The University of Tokyo

Short Paper

**Keywords:** Japanese Buddhism, Jodo Shinshu, Historical Documents

## Abstract

In this study, we embarked on the task of digitizing the "Kyōmyo-chō" using the Text Encoding Initiative (TEI) markup language. The "Kyōmyo-chō" is a pivotal register of the disciples of Shinran, who was the founder of the Jōdo Shinshū sect of Japanese Buddhism. It chronicles the disciples' names, residences, and their master-disciple relationships, offering a myriad of unique personal names absent from other sources. This makes it an indispensable foundation for Shinshū studies and for the digitization of materials related to Shinshū Buddhism.

The "Kyōmyo-chō" adopts a genealogical format that traces the lineage from Shinran to his disciples and grand-disciples. This lineage is depicted in the figure below.

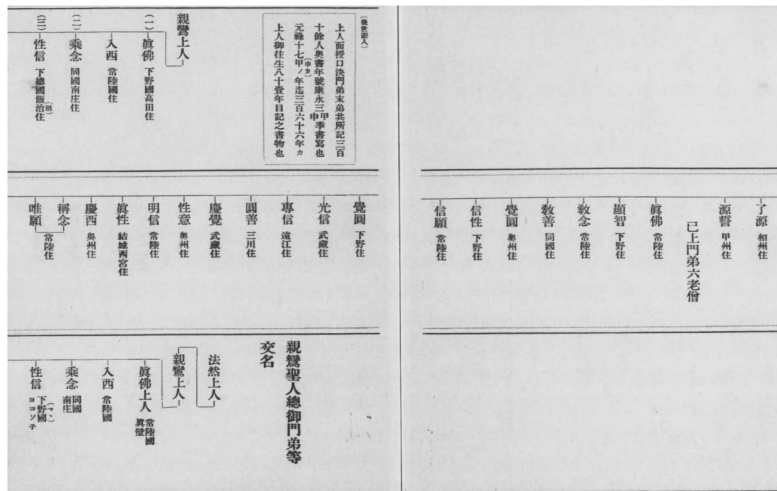


Fig. 1: From the edition of Yamada Bunsho in 1917 (<https://dl.ndl.go.jp/pid/932139/1/7>). Three manuscripts are contrasted in the upper, middle and lower sections, with later

additions enclosed in dotted lines.

Six old manuscripts of "Kyōmyo-chō" exist, raising the question of the optimal text for digitization. We addressed this by encoding the version compiled by Yamada Bunsho in 1917, which integrates three key manuscripts commonly cited in contemporary studies.

We selected Yamada's 1917 edition due to its absence of copyright restrictions, and its frequent citation in related literature. Although there are other noteworthy editions, their accessibility is limited due to copyright constraints. The three manuscripts incorporated into Yamada's 1917 edition present various distinct assertions. The primary source of these differences is the later additions by the scribes, who were attempting to tie their own traditions to Shinran. Additionally, minor discrepancies due to scribal errors also contribute to the variations among the manuscripts. Yamada's 1917 edition demarcates these later insertions with dotted lines, and provides proposed corrections to the scribal inaccuracies found in the manuscripts.

During the markup process of the three texts in Yamada's 1917 edition, we created separate files for each text. This approach ensured the preservation of their unique interpretations of disciple relationships, influenced in part by various additions. To maintain these unique perspectives, we avoided the use of the parallel-segmentation method.

We adhered to a basic approach in marking up personal names using the `<listPerson>` and `<persName>` tags and each master-disciple relationship using the `<listRelation>` and `<relation>` tags. These `<listRelations>`, convertible into RDF format, can then be visualized as a knowledge graph. When a disciple's geographical activity was mentioned, we utilized the `<listPlace>`, `<placeName>`, `<location>`, and `geo` tags to register the location information. In ambiguous cases, for instance, when the place name is too broad or the current location of the place name cannot be identified, we used the `@cert` attribute for provisional adjustments. Additionally, we structured the additional content contained in the manuscripts using `<div>` tags. By doing so, we can determine whether a certain name included in these sections was added at a later time.

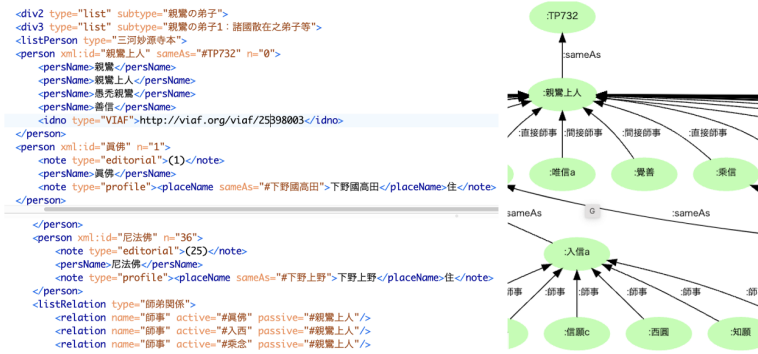


Fig. 2, 3: On the left is a file marked up with listPerson and listRelation tags. The right figure shows an example of a knowledge graph. For Shinran (親鸞上人), we refer to the ID (TP732) of the digital Hobogirin data (<https://tripitaka.l.u-tokyo.ac.jp/hbgrn/Resources>)

Simultaneously, we created a unified file for cross-referencing individual names during the markup process of the three files. This file incorporates all the names from the three manuscripts. The person names in the unified file are referenced to the same person in the three aforementioned files by `<@corresp>` attribute as follows.

```

<person corresp="shinran_jodan.xml#親鸞上人 shinran_chudan.xml#親鸞上人 shinran_
<persName>親鸞</persName>
<persName>愚禿親鸞</persName>
<persName>善信</persName>
<idno type="VIAF">http://viaf.org/viaf/25398003</idno>
</person>
<person corresp="shinran_chudan.xml#明光" n="1" xml:id="明光">
<persName>明光</persName>
<note type="profile"><placeName sameAs="#相模">相模</placeName>住</note>
</person>

```

Fig. 4.

Whenever a disciple's name appears in other documents (e.g., Shinran's letters), we can annotate it and link it to the `xml:id` in this unified file. This approach allows us to reference the mentor-disciple relationships and geographical activities denoted by each manuscript. Manuscripts not included in Yamada's edition can be similarly marked up and added to this unified file at a later stage.

It is noteworthy that many names featured in the "Kyōmyō-chō" lack external resources such as Wikidata. Future endeavors could involve creating Wikidata entries for these significant names and linking them to the unified reference file used in this research. Such a step would significantly enhance the accessibility and visibility of this crucial historical data.

## About the authors

**Yoshihiro Sato** is a Ph.D. student and a project researcher at the University of Tokyo. His main interest is Indian Buddhism and the biography of the Buddha. And he is also interested in the digitization of Jodo Shinshu scriptures. He is also involved in the TEI encoding of Buddhist Chinese texts in the SAT database under Prof. Shimoda.

**Kiyonori Nagasaki**, Ph.D., is a senior fellow at the International Institute for Digital Humanities in Tokyo. His main research interest is developing digital frameworks for collaboration in Buddhist studies. He is also investigating the significance of digital methodology in the Humanities and promoting DH activities in Japan.

**Masahiro Shimoda**, Ph.D., is a former professor at the University of Tokyo, and a professor at Musashino University. He studies the formation process of Mahayana Buddhist scriptures and is also leading a project to build and develop the database of Buddhist Chinese texts (SAT).

# Do ~~not~~ touch the `<object>`. Extending the `<object>` element based on the example of the Visual Archive Southeastern Europe

Selina Galka, University of Graz, Austria

Martina Scholger, University of Graz

Suzana Sagadin, University of Graz, Austria

Long Paper

**Keywords:** visual archive, tei:object, LIDO,  
Southeastern Europe

In its current version, the portal VASE (*Visual Archive Southeastern Europe*, VASE 2012–2023) gathers historical and contemporary visual material from Southeastern Europe in four region-related sub-projects, focusing on the image as a primary source (Derler 2020). The project has been implemented since 2012 at the Centre for Information Modelling of the University of Graz in collaboration with the Centre for South-Eastern European Studies at the University of Graz and with the University of Basel. Since spring 2022, the portal has been undergoing a comprehensive redesign in cooperation with the Austrian Academy of Sciences and the Center for the Study of Balkan Societies and Cultures. The redesign covers the data model, the data curation and the interface.

This paper deals with the question of appropriate modelling of visual material (e.g. postcards and photographs), which is provided with extensive metadata and textual descriptions and can also be a text carrier itself. From the beginning of the project, the development of the data model was characterized by this dichotomy of visual material and text-based descriptions, which ultimately led to the consideration of both the TEI (Text Encoding Initiative) standard and LIDO (Lightweight Information Describing Objects) – an XML schema for encoding museum and collection objects (ICOM 2010) – for data modelling. Three of the portal’s projects have been implemented with TEI to allow for deeper indexing of textual descriptions and commentaries, which is not possible in LIDO.

With the introduction of the `<object>` element in TEI, this shortcoming could now be addressed: In the course of the redesign, the previously semantically weak TEI



model, which consisted primarily of typified <ab> elements, was replaced by the semantically more expressive <object> structure. The newly developed data model is based on the TEI element <object>, which was introduced in version 3.5.0 of the TEI P5 Guidelines to describe a “single identifiable physical object” (TEI Consortium 2019). So far, the element basically contains the same content that is available for manuscript description within <msDesc>. Thus, within <objectIdentifier>, country (<country>), place (<placeName>), repository (<repository>) and inventory number (<idno>) are specified. In <physDesc>, the object type (<objectType>) and material (<material>) are encoded within <p> elements. The <objectDesc> contains information on the form (@form, e. g. “Carte de visite”) and dimensions (<supportDesc>) of the visual material.

However, the elements permitted so far within <object> do not yet reflect all our needs. For example, the <objectContents> element (analogous to <msContents>) was defined within the project and contains text-based object descriptions or comments within <ab>.

On the one hand, the paper presents a proposal for an extension of the <objectContents> element beyond mirroring elements from the <msDesc> element, which – drawing on the experience with modelling visual sources gained in this project – will be submitted to the TEI Technical Council for review. This proposal also considers Nelson’s (2017) assessment of the potential for generalizing existing TEI elements describing manuscripts, e.g. <material>, <desc>, <dimensions>, or <objectType>, to make them applicable to objects of any kind. On the other hand, it shows the possibilities of linking text-centric TEI-based modelling with concepts of the object-centric XML standard LIDO in order to establish compatibility with this standard for subject-specific use cases such as the modelling of text-bearing visual sources in continuation of efforts to map the TEI with conceptual models (e.g CIDOC CRM) from the domain of cultural heritage (Ore and Eide 2009).

## Bibliography

Derler, Barbara. 2020. “Südosteuropaforschung delokalisiert & digital. Das Visuelle Archiv Südöstliches Europa (VASE)”. In *From the Highlands to Hollywood*, edited by Siegfried Gruber, Dominik Gutmeyr, Sabine Jesner, Elife Krasniqi, Robert Pichler, Christian Promitzer, 163-180. Wien: LIT.

ICOM. International Committee for Documentation. 2010. "Lightweight Information Describing Objects – LIDO." [http://network.icom.museum/cidoc/working-groups/lido/lido-technical/specificati%C2%ADon/ <http://network.icom.museum/cidoc/working-groups/lido/lido-technical/specification/>](http://network.icom.museum/cidoc/working-groups/lido/lido-technical/specificati%C2%ADon/http://network.icom.museum/cidoc/working-groups/lido/lido-technical/specification/).

Nelson, Brent. 2017. "Curating Object-Oriented Collections Using the TEI." *Journal of the Text Encoding Initiative*, Issue 9 (September 2016 - December 2017). <https://doi.org/10.4000/jtei.1680>.

Ore, Christian-Emil, and Øyvind Eide. 2009. "TEI and cultural heritage ontologies: Exchange of information?" *Literary & Linguistic Computings* 24 (2): 161–72. <https://doi.org/10.1093/lc/fqp010>.

TEI Consortium, eds. 2019. "<object>" *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 3.5.0. <https://tei-c.org/Vault/P5/3.5.0/doc/tei-p5-doc/en/html/ref-object.html>.

VASE. Visual Archive Southeastern Europe. 2012–2023. <https://gams.uni-graz.at/vase>.

# Encoding in human centered machine learning workflows: case study on mensural ligature recognition

**David Rizo**, Universidad de Alicante, Spain; Instituto Superior de Enseñanzas Artísticas de la Comunidad Valenciana (ISEA.CV), Spain

**Jorge Calvo-Zaragoza**, Universidad de Alicante, Spain

**Juan Carlos Martínez Sevilla**, Universidad de Alicante, Spain

**Antonio Madueño**, Universidad de Alicante, Spain

**Patricia García-Iasci**, Universidad de Alicante, Spain; Universidad de Salamanca, Spain

**Teresa Delgado-Sánchez**, Biblioteca Nacional de España (BNE), Spain

**Short Paper**

**Keywords:** OMR, encoding, UX, machine learning, transcription

## Abstract

This work presents a case study on the human-centered design of machine learning workflows for Optical Music Recognition (OMR). The study specifically targets mensural ligature recognition, which presents unique challenges for OMR due to its complex notation and variations in shapes. The case study is conducted within the context of a larger project that aims to encode all books of mensural notation held by the National Library of Spain (BNE) [\[1\]](#) using the online version (Rizo et al. 2023) of the Music Recognition, Encoding, and Transcription Tool (MuRET) (Rizo et al. 2018).

The study employs an incremental approach that builds on a seed model created from an existing training collection (Parada-Cabaleiro et al. 2019). The output of the OMR is post-edited to obtain a high-quality encoding, and the corrected sources are used to create new models that improve OMR accuracy and reduce post-editing time.

In (Calvo-Zaragoza and Rizo, 2018), the differentiation between *agnostic* and *semantic* content was introduced to reduce the variability of symbols the recognition systems had to detect, and to allow automatic transcriptions to be

corrected by non expert editors who should only be able to identify graphical symbols. This approach has been extensively used in the work that has enabled to build a big enough corpus to train high quality OMR deep learning models.

In the transcription process assisted by OMR, the most difficult challenge has been the transcription of ligatures in which this paper will focus. Initially, the models were unable to learn how to recognize ligatures because of their scarce occurrence, so they had to be represented graphically as a generic "ligature" symbol, and semantically encoded in **\*\*mens** (Rizo et al. 2019). This manual process significantly slowed down the whole post-editing process, making it more time-consuming and labor-intensive.

A synthetic corpus was built to experiment with different agnostic encodings in order to select those ones with the best trade off between training corpus size and performance. Among those encodings, experiments with real end users were carried out to find out the more user-friendly agnostic representation in terms of user experience (UX) for the process of fixing classification errors.

Once a considerable number of manually encoded ligatures were obtained in all the real corpora in MuRET using the generic ligature symbol for the agnostic representation, and semantically encoded using **\*\*mens**, they were converted to the new agnostic encoding using a state machine transducer to obtain the final representation, from which new models were created that correctly recognized the ligature in the sources.

Thanks to this approach, the recognition and post-editing effort were reduced by a factor of 10, allowing the end user to obtain a complete and correct encoding of a standard book page in under one minute per page.

## Bibliography

Calvo-Zaragoza, J. and David Rizo. 2018. "End-to-End Neural Optical Music Recognition of Monophonic Scores" *Applied Sciences* 8, no. 4: 606.

<https://doi.org/10.3390/app8040606>

Parada-Cabaleiro, E., Batliner, A., & Schuller, B. W. (2019). "A Diplomatic Edition of Il Lauro Secco: Ground Truth for OMR of White Mensural Notation". In A. Flexer, G. Peeters, J. Urbano, & A. Volk (Eds.), *Proceedings of the 20th International Society for Music Information Retrieval Conference, ISMIR 2019, Delft, The Netherlands, November 4-8, 2019*, pp. 557-564.

Rizo, D., Calvo-Zaragoza, J., & Iñesta, J. M. (2018). "MuRET: a music recognition, encoding, and transcription tool". In K. R. Page (Ed.), *Proceedings of the 5th International Conference on Digital Libraries for Musicology, DLfM 2018, Paris, France, September 28, 2018*, pp. 52–56. ACM.

<https://doi.org/10.1145/3273024.3273029>

Rizo Valero, D., Pascual León, N., & Sapp, C. S. (2019). "White Mensural Manual Encoding: from Humdrum to MEI", *Cuadernos De Investigación Musical*, (6), 373–393. <https://doi.org/10.18239/invesmusic.v0i6.1953>

Rizo, D., Calvo-Zaragoza, J., Martínez-Sevilla J.C., Roselló, A., and Fuentes-Martínez, E. (2023). "Design of a music recognition, encoding, and transcription online tool", *16th International Symposium on Computer Music Multidisciplinary Research, Tokyo, 2023* (accepted)

## Notes

---

1. <https://grfia.dlsi.ua.es/polifonia/> (accessed April 1st, 2023) ↵

# Encoding mathematics: A concept for a data model for representing early modern and modern mathematical notations in Digital Editions

**Elisabeth Rinner**, Leibniz-Forschungsstelle Hannover der Niedersächsischen Akademie der Wissenschaften zu Göttingen beim Leibniz-Archiv der Gottfried Wilhelm Leibniz Bibliothek – Niedersächsische Landesbibliothek Hannover (GWLB/NAdWGö), Germany

**Long Paper**

**Keywords:** historical mathematical notations, multi-modal texts, digital editions, data model, linked data

## Abstract

In my paper, I will present a concept for encoding historical mathematical texts. As this work is part of an investigation of possibilities to expand the use of digital methods in the *Leibniz Edition*, it is based on passages from mathematical writings of G.W. Leibniz ([LSB1923ff]).

In 17th centuries' European mathematics, a notation system for writing down mathematical content developed. Nevertheless, specific notations and spatial arrangements are also used in procedures such as long division, or they provide means that aim on establishing new knowledge. By this, notations are an essential part of "doing mathematics". As notations and their specific characteristics vary between scholars and over time, they are object of investigation themselves, and preserving them in text editions is crucial for research in the history of mathematics ([Unguru1975]).

However, [TEI P5] proposes to use markup languages which do not "tag" the text according to the actual structures of historical notations—if they can represent them at all. Since an adequate encoding is missing, digital editions of such texts are rare, and existing specimen can scarcely be evaluated digitally in a way that is meaningful for historians of mathematics ([Briefportal], [Görmar2023], [Newton]).

The encoding is indeed challenging: the data model needs to be capable to represent deviations from linear text flow as well as complex and multiple spatial relationships between text elements. In complex spatial arrangements, even the meaning of a single letter can depend on its context if it is part of several interwoven text elements. This illustrates that mathematical notations exceed the limits of today's manifestation of TEI in [TEI P5] (cf. e.g. [Sahle2013] p. 374). In addition, some passages prove to be multi-modal, in that in one case it is the text strings, and in others rather the spatial arrangements from which meaning is derived ([Rinner2023], section "Résumé und Ausblick").

The core idea of the proposed approach is to regard mathematical notations as a sign system in the sense of Ch.S. Peirce's semiotics (so the signs are characterized, among others, by their meaning and their visual representation), and to individuate their structure elements accordingly. A documentation of the signs can be a resource for research on its own. In order to connect several digital editions, a collaboratively established "Dictionary of mathematical notations" seems to be a desideratum for the history of mathematics.

Representing the semiotic signs as nodes in a graph-based approach leads to an appropriate encoding of the complexities of mathematical notation. By separating the "bits" of mathematical notations such as short text strings (often, they will comprise only single characters) and additional graphical elements from the encoding of structure elements, re-interpretations as they can be observed within and across historical mathematical texts as well as additional research-specific perspectives of mathematical notation can be added to an already existing digital edition. By capturing different modes such as textual or diagrammatic aspects the text is represented in full depth, and options for searching it can be expanded.

A useful and well-adjusted set of components for an implementation is still to be established. Due to the character of this approach, this should be preceded by discussions in the history of mathematics.

## Bibliography

[Briefportal] [Leibniz-Forschungsstelle Hannover der Akademie der Wissenschaften zu Göttingen beim Leibniz-Archiv der Gottfried Wilhelm Leibniz Bibliothek – Niedersächsische Landesbibliothek]: *Briefportal Leibniz. Ausgewählte Briefe in HTML*. <https://leibniz-briefportal.adw-goe.de/>.

[Görmar2023] Maximilian Görmar: *Briefportal Leibniz. Ausgewählte Briefe in HTML*, Niedersächsische Akademie der Wissenschaften zu Göttingen in Kooperation mit der SUB Göttingen (ed.), 2016-2019. (Last Accessed: 29.08.2022). Reviewed by Maximilian Görmar (Herzog August Bibliothek Wolfenbüttel). In: *RIDE – A review journal for digital editions and resources, Issue 16: Scholarly Editions (FAIR criteria)*, ed. by Tessa Gengnagel, Frederike Neuber, Daniela Schulz and Ulrike Henny-Krahmer, Frederike Neuber, Martina Scholger as managing editors, 2023. DOI: [10.18716/ride.a.16.3](https://doi.org/10.18716/ride.a.16.3).

[LSB1923ff] Gottfried Wilhelm Leibniz: *Sämtliche Schriften und Briefe*, ed. by Preußische Akademie der Wissenschaften et al., Darmstadt et al. 1923ff. A substantial part of volumes as well as additional material to the Leibniz Edition such as pre-print versions of upcoming volumes and comprehensive registers can be accessed online via <https://leibnizedition.de/>, <https://www.gwlb.de/leibniz/digitale-ressourcen/repositorium-des-leibniz-archivs/> and <https://rep.adw-goe.de/handle/11858/00-001S-0000-0006-AB16-C>.

[Newton] *The Newton Project*, <https://www.newtonproject.ox.ac.uk>.

[Rinner2023] Elisabeth Rinner: *Die Zerfallungsschemata der „Regula discriptionum et triscerptionum universalis“ von Gottfried Wilhelm Leibniz*. In: [Tagungsband der gemeinsamen Jahrestagung der Fachsektion „Geschichte der Mathematik“ der DMV und des Arbeitskreises „Mathematikgeschichte und Unterricht“ der GDM vom 15.03.2023 bis 19.03.2023 in Naunhof bei Leipzig (preliminary title)] (submitted).

[Sahle2013] Patrick Sahle: *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). BoD, Norderstedt, 2013.

[TEI P5] The TEI Consortium: *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Originally edited by C.M. Sperberg-McQueen and Lou Burnard for the ACH-ALLC-ACL Text Encoding Initiative now entirely revised and expanded under the supervision of the Technical Council of the TEI Consortium. Version 4.5.0. Last updated on 25th October 2022, revision 3e98e619e. Text Encoding Initiative Consortium, 2022. <https://tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf> (accessed on 04/04/2023).

[Unguru1975] Sabetai Unguru: *On the Need to Rewrite the History of Greek Mathematics*. In: *Archive for History of Exact Sciences*. Vol. 15, No. 1 (1975), pp. 67-



114. DOI: [10.1007/bf00327233](https://doi.org/10.1007/bf00327233)

## About the Author

**Dr. Elisabeth Rinner**, DH person and editor of Leibniz's mathematical writings in the project *Leibniz Edition* at the *Leibniz-Archiv* (= *Leibniz-Forschungsstelle Hannover der Niedersächsischen Akademie der Wissenschaften zu Göttingen beim Leibniz-Archiv der Gottfried Wilhelm Leibniz Bibliothek – Niedersächsische Landesbibliothek* (GWLBNAdWGö) in Hanover/Germany, <https://www.gwlb.de/leibniz/leibniz-archiv>). Research interests include data models, digital publishing, methodology of DH, and history of mathematics (especially mathematical notations).

# Encoding Notre Dame polyphony for corpus analysis

Joshua Stutter, University of Glasgow, United Kingdom

Long Paper

**Keywords:** Early music, Notre Dame, Historiography, Directed graphs, Corpora

## Abstract

The encoding of 13th-century vocal music, known as Notre Dame (ND) polyphony, presents novel challenges for music encoding. Beyond the peculiarities of its notation, including unwritten notes, diverging variation in sources, and a subtle interplay of musical reuse between the two types of notations for organa and motet, this paper argues that the main challenges facing a satisfactory encoding of ND polyphony are: a) the representation of modal rhythm, and b) a lack of academic consensus concerning certain basic aspects of the notation.

Unlike unmeasured monophony such as plainchant or fully rhythmically-specified polyphony such as mensural notation or CWMN, an encoding of ND polyphony cannot be structured by voice leading or duration. The highly idiosyncratic and contextual ligature patterns of ND notation can seldom be interpreted using the system of rhythmic modes: it requires the rules to be bent or broken or, in most cases, left to the discretion of a performer. As a result, the alignment of voices becomes difficult to encode and there are numerous “correct” answers. The only clues are in the differing graphical layouts of the scribes.

Moreover, despite a century of musicological research into the ND repertory, the academic community has yet to reach a common consensus on some of its most fundamental issues. This is due, in part, to the paucity of surviving sources and contemporary writing, the influence of oral traditions on the transmission of sources, and the repertory's exploitation by the shifting narratives of music history as the so-called “crucible” of Western art music. Typical approaches to encoding ND polyphony must therefore shoehorn a transitional style into either a “neumatic” or “mensural” context when it in fact belongs to neither.

To address these challenges, this paper proposes a new approach to encoding ND polyphony as part of the “Clausula Archive of the Notre Dame Repertory” (CANDR). The approach is based on a two-layer directed graph that avoids neumatic or mensural structures. By expanding the graph into a complete graph using a simple distance metric, this paper demonstrates the building of node embeddings and continuous-bag-of-words embeddings to vectorize sections of music, as well as the training of a support vector machine (SVM) to extract passages of interest. This approach offers a viable solution for addressing the challenges presented by this unique repertory.

# Encoding Orchestral Parts that Share a Staff in MEI: Guidelines and a Parts Extraction Tool

Oleksii Sapov-Erlinger, Mozarteum Foundation Salzburg, Austria

Short Paper

**Keywords:** digital music edition, MEI, music notation engraving, score parts

## Abstract

It is common engraving practice for “two or more players – or a divided string section with only single notes per division” (Gould 2011) to share a staff in an orchestral score, such as clarinet I and clarinet II. This practice addresses issues such as stemming, sharing musical content, placement of ties and slurs, and more. With respect to these engraving conventions, a comprehensive music encoding should also capture the semantic independence of the parts, which is essential for parts extraction or performing automated analysis.

Currently, there are no exhaustive encoding guidelines that adequately address this topic. MEI provides a few examples and general recommendations, but there is no dedicated section such as *Stave sharing* (cp. Gould 2011). Additionally, there are no encoding guidelines from other projects or relevant publications. Only a few related MEI encodings are available in open access, but they tend to have a bias towards the visual domain.

The proposed encoding guidelines were developed in the context of the [Digital Interactive Mozart Edition](#) for the encoding of the symphonic repertoire. The guidelines establish general rules and offer case-based solutions, as well extending MEI. The fundamental ideas of the guidelines are as follows: 1) consistent separation of Events onto layers, 2) usage of the `@sameas` attribute for shared Events, and 3) usage of the `@layer` attribute for an unambiguous assignment of ControlEvents to Events. These rules, along with extensions for specific notation cases, will be explained in detail in the paper. For MEI rendering, the Verovio engraving library is utilized, benefiting from suggestions and improvements made during the creation of the guidelines.

As a proof-of-concept, a basic parts extraction tool has been developed with the potential to evolve into a full-fledged MEI parts creation tool. *Extract Parts* has been published on GitHub (<https://github.com/ism-dme/DIME-tools>) and is available for use by any project, provided that the MEI data adheres to the encoding guidelines proposed in this paper.

## Bibliography

*Digital musical analysis with MEI on the example of Anton Bruckners' compositional studies*. 2019. *The Bruckner Study Book Viewer*.

Version: a7e2660. <https://github.com/paavomare/Bruckner-Study-Book-Viewer.git>

*Freischuetz-Digital*. 2017. "data-music." Version: v1.1.0.

<https://github.com/Freischuetz-Digital/data-music.git>

Gould, Elaine. 2011. *Behind bars: The definitive guide to music notation*. London: Faber Music.

Music Encoding Initiative. 2019a. "MEI Guidelines."

Version: v4.0.1. <https://music-encoding.org/guidelines/v4/content>

Music Encoding Initiative. 2019b. "The MEI Sample Encodings." Version: 4ae1149.

<https://github.com/music-encoding/sample-encodings>

Music Encoding Initiative. 2023. "MEI Guidelines." Version: v5.0.0-dev, cef8ce1.

<https://music-encoding.org/guidelines/dev/content>

Pugin, Laurent, Rodolfo Zitellini, and Perry Roland. 2014. "Verovio: A library for Engraving MEI Music Notation into SVG." In *Proceedings of the 15th International Society for Music Information Retrieval Conference*, 107–112, Taipei, Taiwan. Taipei, Taiwan: ISMIR, October. <https://doi.org/10.5281/zenodo.1417589>

## About the author

**Oleksii Sapov-Erlinger**, MA, ([0009-0003-7012-5781](tel:0009-0003-7012-5781)), has been working since 2018 as a research assistant at the [Digital Interactive Mozart Edition](#), a joint project of the *Mozarteum Foundation Salzburg* and the *Packard Humanities Institute*, with focus on Data Modeling and Data Engineering in MEI. Additionally, since July 2022, he has been working as a research assistant at the [Die Wenzelsbibel – Digitale Edition](#)

[und Analyse](#), a joint project of the University of Salzburg and the Austrian National Library. His focus in this position is on TEI Data Engineering and Web Development.

# Encoding previously Transcribed Oral Literature: The Forager Folklore Database (FFDB)

Jan Niklas Jokisch, Max Planck Institute for Empirical Aesthetics, Germany

Poster

**Keywords:** folklore studies, narrative theory, general and comparative literature, ethnology, anthropology

## Abstract

The Forager Folklore Database (FFDB) assembles a large corpus of hunter-gatherer oral literature in English translation. It is designed as an empirical basis for the systematic study of narrative universals and the evolutionary origin of storytelling, as well as for (comparative) folklore studies.

The folktales in question were originally recorded by anthropologists and others (clergy, explorers, etc.) during the past 150 years. The FFDB, aside from generating rich metadata for each narrative indicating its provenance as well as selected textual features, will provide a corpus of narratives as digital texts encoded in TEI XML. The encoding preserves spelling and page beginnings of the original source, thus making the oftentimes hard to access material available in a citable format. Furthermore, semantic annotation is set up for animals and plants (semi-automatically through WordNet synsets), colors, person and place names as well as for narrative categories such as ana- and epimythia. This will offer researchers a more fine-grained method of text retrieval, e.g., finding all texts containing birds, and allow for richer computerized approaches in the growing field of computational folktale and narrative studies.

The form of previously transcribed orality that the folktales take comes with its own set of challenges. Though they are transmitted to us in writing, many of the relevant features of the narratives derive from their origin in oral storytelling.

Conversational framings, explanatory comments, personal asides, remarks and questions from the audience, the use of gestures, onomatopoeia, and songs go beyond the generally more bookish approaches of the TEI. They require new tools, new tags, and a new understanding of textuality and authorship.

The general project infrastructure is set up dynamically in the form of a work-in-progress relational database. We use Python to generate the TEI header automatically from the data stored in the database, before combining the headers with the pre-annotated texts. Unless more restrictive copyright prohibits this, the files will be published under a CC BY-NC-SA license. The metadata for the encoded narratives are enriched further by the motif assignments in Stith Thompson's *Motif-Index of Folk-Literature* (1950-58), together with additional motif assignments, e.g., by Johannes Wilbert and Karin Simoneau (*Folk Literature of the South American Indians*, 1970-1992) and Sigrid Schmidt (*Catalogue of the Khoisan Folktales of Southern Africa*, 2013). A small selection of roughly 70 narratives has been encoded already to showcase the direction and possible scope of the FFDB.



# Encoding Traditional Court Music and Dance Scores Using the Music Encoding Initiative

Shintaro Seki, The University of Tokyo, Japan

Long Paper

**Keywords:** Gagaku, Traditional Japanese Music, Asian Sheet Music

## Abstract

Gagaku, a comprehensive art form encompassing instrumental music, vocal music, and dance, originated during the mid-Heian period (around the 10th century) and has been transmitted in its fundamental form to the present. In addition to its role as court music, gagaku offers numerous performance opportunities at shrines and temples. Its repertoire not only includes ancient Japanese music but also embraces the musical cultures of Mainland China and the Korean Peninsula. Consequently, gagaku occupies a significant position in the overall musical culture of East Asia, preserving some of the music that has been lost in China and Korea<sup>[1]</sup>.

Like other non-Western musical cultures, gagaku faces significant challenges related to digital archiving, encoding, and accessibility<sup>[2]</sup>. The dissipation of original documents poses a critical problem, endangering valuable resources. Moreover, the limited number of sheet music users hampers the digitization of materials. This paper aims to preserve richer information and enhance the accessibility of gagaku scores by encoding them in machine-readable form using the Music Encoding Initiative (MEI).

Gagaku scores are typically written vertically in Kanji, Katakana, and other special symbols prepared for gagaku scores, making it challenging to transcribe them into Western staff notation, commonly used in music informatics and music information retrieval. While previous studies have attempted to convert gagaku notation into staff notation, most have focused solely on converting actual gagaku performances into staff notation, neglecting the structure of gagaku notation. Consequently, these studies concentrate on gagaku performances rather than gagaku scores. Furthermore, gagaku scores primarily serve to support the performer's memory and

are not intended to facilitate the dissemination of unknown music, as staff notation does. Thus, many descriptions in gagaku scores are ambiguous, assuming that the musician has learned the piece through lessons.

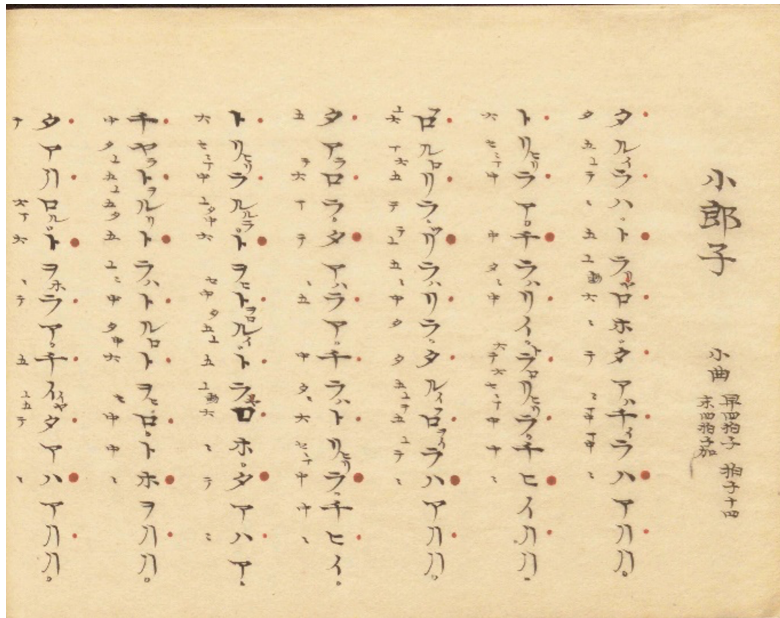


Fig. 1: Examples of Gagaku Score for Shō (Available at <https://kokusho.nijl.ac.jp/biblio/100376839/717>)

One of the main features of gagaku scores is the presence of “shōga.” Shōga consists of words specifically prepared for learning the melodic phrases, enabling learners to acquire the correct melody and rhythm of a piece of music by repeatedly singing shōga after the teacher. Although it can be considered a mnemonic, shōga itself lacks any linguistic meaning and is purely intended for memorizing the melody. As such information does not exist in other notation systems, including staff notation, it is essential to consider appropriate encoding methods.

MEI, an XML-based music encoding guideline, offers a flexible descriptive framework for non-Western music scores, allowing accurate encoding of gagaku scores that often contain abstract oral descriptions based on oral traditions. This paper will discuss the transcription of three main instruments (Shō, Hichiriki, and Ryūteki) in gagaku ensembles using MEI. gagaku encompasses instrumental, vocal, and dance music, each described using a different notation system. The study will focus on music belonging to the genre called “tōgaku,” comprising pieces derived from the Tang Dynasty. Tōgaku exhibits a relatively well-defined cyclic rhythm among gagaku pieces, making its structure closer to Western music. Therefore,

Tōgaku serves as an appropriate target for an initial test of gagaku score encoding, with a focus on the special description system of gagaku scores.

Additionally, this paper explores the possibility of new Japanese music research that effectively leverages research findings from computer science and digital humanities, applying the results of digital musicology and music information retrieval to East Asian musical score culture.

## Bibliography

### Acknowledgement

This work was supported by Foundation for Cultural Heritage and Art Research.

### About the author

**Shintaro Seki** is a Japan Society for the Promotion of Science Research Fellow DC2 and a Ph.D. student in the Graduate School of Humanities and Sociology, The University of Tokyo. His research aims to promote the study of Japanese music through an informatics-based approach.

- 
1. Tokita, Alison McQueen, and David W. Hughes, eds. 2008. *The Ashgate Research Companion to Japanese Music*. London: Routledge. [↩](#)
  2. Lidy, Thomas et al. 2010. "On the Suitability of State-of-the-Art Music Information Retrieval Methods for Analyzing, Categorizing and Accessing Non-Western and Ethnic Music Collections." *Signal Processing* 90(4): 1032–48. [↩](#)

# Encoding Transcultural Texts - Applying TEI to Early Modern Chinese Christian Literature

Wenlu Wang, The University of Tokyo, Japan

Poster

**Keywords:** Christian literature, transcultural texts, translation, data modeling

## Abstract

This project centers on data modeling and encoding of early modern Chinese Christian literature. It demonstrates the effectiveness of structuring and encoding these texts in TEI-XML for accumulating and analyzing Chinese Christian text data, as well as contributes to the discussion on developing models for encoding and linking transcultural texts.

Over 1,050 works were co-produced by European missionaries and Chinese converts between 1582 and the mid-19th century. These texts, including religious manuals and translations of Scholastic philosophy, document various aspects of the transcultural encounters between Europe and China during the early modern period. Consequently, they serve as essential source materials for research on, among other topics, the Catholic Church's global mission and global knowledge circulation in the early modern era. In the past few decades, bibliographic databases and large-scale digitization and reprint projects of collections in European and Chinese libraries and archives have made a major part of the texts available to the research community. Thanks to these publications, a series of print-based transcriptions have been published, further improving the accessibility of the texts.

However, digital versions with searchable full-text are not yet available; moreover, no efforts have been made to provide the text data in structured machine-readable formats, such as TEI-XML. Drawing from experiences of encoding text in East Asian languages in TEI-XML, especially Buddhist texts, this project applies TEI encoding to Chinese Christian texts, exploring possibilities of making the texts available in more economical, sustainable ways and to a wider community.

As an initial effort, this project encoded a collection of texts known as "Dottrina Christiana," which consists of fundamental Christian tenets that converts need to learn before receiving baptism and recite in religious practices afterwards. Often anonymous, this genre contains dozens of variants that differ in length and specific word usage, which previously made comprehensive analysis challenging and unsuccessful. The critical apparatus framework of the TEI guidelines proved to be effective in recording the variants in an efficient and precise manner while allowing future additions when a new variant is identified. The project adopts the parallel segmentation method and encodes according to the neutral style (Figure 1), as the purpose of the project is to provide an overall understanding of the fluidity of the texts and evidence for grouping and dating each variant. Besides the `<app>` element, the encoding also utilizes a set of structural tags, including `<front>`, `<body>`, `<back>`, and under the `<body>` element, `<div>`, `<head>`, `<p>` to show the basic structure of the text.

```
<body>
  <div n="1" xml:id="PaterNoster"
    type="section">
    <head>天主經</head>
    <p>在天我等父者。我等願爾名 <app>
      <rdg wit="#a_ARSI126">成聖 </rdg>
      <rdg
        wit="#b_BAV221_5 #d_BAV132_1 #c_BNF7373 #f_BNF6972 #g_BAV285_7"
        >見聖</rdg>
      <rdg wit="#e_ARSI57">丕顯</rdg>
    </app>。
```

Fig. 1.

At this point, the project has encoded seven primary variants and utilized existing visualization tools such as Versioning Machine and Critical Apparatus Toolbox alongside to revise and improve the encoding. Based on the encoded data and visualizations, this project managed to provide the first systematic analysis of this genre, grouping and dating the variants. This enables mapping the gradual development of translations of key theological terminologies and contents to teach the catechumens and converts at different time periods. Although both the encoding guidelines and visualization tools generally worked well with these early modern Chinese Christian texts, several issues remain unresolved. Among the challenges are how to encode and visualize *wari-gaki* 割書 (*shuanghang xiaozhu* 双行小注 in Chinese, i.e. texts written in smaller size in two equally split lines, functioning as annotations to the main text, Figure 2), and how to present and compare each section encoded as `<div>`, that varies in length, in a parallel fashion, which the project couldn't realize using existing visualization tools.

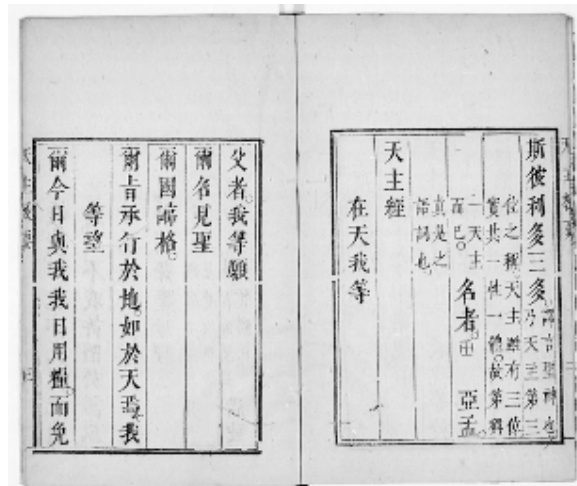


Fig. 2.

Chinese Christian texts are transcultural in nature. Some are direct translations from works in European languages, and others are free adaptations. They are also part of the multilingual body of works produced in the early modern global mission of the Catholic Church, whose activities ranged from Asia, the Americas, through Africa. Drawing from previous studies of parallel corpora of Buddhist texts and multilingual Bibles, this project is exploring ways to link Chinese Christian texts with their European origins and early modern Japanese Christian texts that are translated from the same European originals. In addition to paragraph or sentence-level parallel corpus, the project also plans to encode key theological terminologies that are rendered into Chinese and Japanese according to either pronunciation or meaning. Encoding Chinese Christian literature as transcultural texts shares general practices and common challenges with multilingual corpus building in Buddhist studies and Biblical studies. The shared practice may serve as the gateway through which researchers of the History of Christianity in Asia can connect with broader research communities.

## Bibliography

Apollon, Daniel, Claire Bélisle, and Régnier Philippe, eds. *Digital Critical Editions*. University of Illinois Press, 2014.

Burghart, Marjorie. "The TEI Critical Apparatus Toolbox: Empowering Textual Scholars through Display, Control, and Comparison Features." *Journal of the Text Encoding Initiative*, no. 10 (December 7, 2016). <https://doi.org/10.4000/jtei.1520>.

Nagasaki, Kiyonori, Toru Tomabechi, and Masahiro Shimoda. "Towards a Digital Research Environment for Buddhist Studies." *Literary and Linguistic Computing* 28, no. 2 (June 2013): 296–300. <https://doi.org/10.1093/lc/fqs076>.

Schreibman, Susan, Amit Kumar, and Jarom McDonald. "The Versioning Machine." *Literary and Linguistic Computing* 18, no. 1 (April 1, 2003): 101–7. <https://doi.org/10.1093/lc/18.1.101>.

Standaert, Nicolas, and Nora Van den Bosch. "Mapping the Printing of Sino-European Intercultural Books in China (1582–c.1823)." *East Asian Publishing and Society* 12, no. 2 (October 11, 2022): 130–91. <https://doi.org/10.1163/22106286-12341367>.

Standaert, Nicolas, and Ad Dudink. "Chinese Christian Texts Database (CCT-Database)." Accessed August 1, 2023. <http://www.arts.kuleuven.be/sinology/cct>.

The CrossWire Bible Society. "OSIS Manual." Accessed August 1, 2023. <https://crosswire.org/osis/>.

The SAT Daizōkyō Text Database Committee (SAT). "The SAT Daizōkyō Text Database." Accessed August 1, 2023. <http://21dzk.l.u-tokyo.ac.jp/SAT/>.

## About the author

**WANG Wenlu** is a project researcher at Tokyo College, The University of Tokyo. Her research field includes the history of Christianity in China, and Europe–East Asia intercultural encounters. She is now working on a manuscript about Chinese translations of Christian doctrine (1580s–1720s) and a TEI critical edition of a Manila Chinese Christian work.

# Exploring the Genesis of Complex Music Manuscripts

Johannes Kepper, Beethovens Werkstatt, Germany

Richard Sanger, Beethovens Werkstatt, Germany

Jan-Peter Voigt, Beethovens Werkstatt, Germany

Long Paper

**Keywords:** digital music edition, genetic criticism, Beethoven, MEI

## Abstract

A common task in humanities is to explore the genesis of textual or musical works. While the creative processes behind a composition are usually not directly accessible to our investigation, we can sometimes deduce them by traces left on scrap paper. For genetic editions, these traces are crucial to understanding the original material relating to a composition, and to identifying as many intermediate steps in the creation of the work(s) in question as possible. Sketchbooks are challenging for a number of reasons. Sketchbooks are personal documents, that normally are not intended for any audience other than the writer. The writing is therefore often rather untidy. This is seconded by the fact that sketches in music are often incomplete as they do not specify all musical parameters found in traditional music notation. For instance, a series of noteheads with stems would usually be interpreted as a sequence of quarter notes. In sketch material, however, the very same scripture could just indicate a series of pitches, without making statements about the rhythm of that series of "notes". In addition, musical sketches often compile music material differently compared to the way in which it is implemented later in the compositional process. For instance, an idea for the section of a full-orchestral work might be sketched in one or two staves only. The visual appearance of such sketches compared to their final implementation in such cases is significantly different, which, as a result, requires an intimate knowledge of the final work to be able to relate the sketch to a corresponding passage in the final work. This is often complicated by the fact that musical sketchbooks may contain material and ideas for multiple works composed more or less simultaneously. Since not every sketch is actually taken up and eventually included in the final composition, this introduces a significant amount of uncertainty, making the



determination of a sketch a task that is far from trivial and requires interpretation on many different levels.

Modelling such material together with its editorial treatment for a digital edition results in significant complexity. Following the basic approach of the TEI's original proposal for an encoding model for genetic editions<sup>[1]</sup>, such a model for music documents makes use of multiple encodings of the same content, gradually moving from the description of the visual appearance of shapes on pages, through their interpretation as musical symbols, the editorial supplement of context to turn such sketches into musically valid texts, and eventually to the relation to finished works. While such a model will help to ensure transparency of editorial work, creation and maintenance of data conforming to this model poses notable challenges because of an inevitably large number of elements linked with each other across multiple encodings.

Our paper will introduce a prototype software called *FacsimileExplorer*, which assists in creating encodings for digital genetic editions of music sketchbooks. It implements a data model based on the combination of SVG and multiple variants of MEI.

## Notes

---

1. c.f. <https://tei-c.org/Vault/TC/tcw19.html> ↵

# FAIR Derived Data in TEI for Copyrighted Texts

**José Calvo Tello**, University of Göttingen, Germany

**Keli Du**, Trier Center for Digital Humanities

**Mathias Göbel**, University of Göttingen, Germany

**Nanette Rißler-Pipka**, Gesellschaft für wissenschaftliche Datenverarbeitung mbH (GWDG)

**Long Paper**

**Keywords:** derived data, FAIR, copyright issues, literature

## Abstract

In this proposal, we present several options for encoding data derived from copyrighted texts in TEI. We argue for the benefits that TEI can bring to the FAIR status of the different types of data (text, textual structure, metadata, annotation), and present each option in five corpora which cover different languages, genres and periods.

## Proposal

Text and data mining using copyrighted texts faces restrictions on publication and re-use. One solution to this problem is to transform the original files to derived data (see e.g. Sánchez Sánchez and Domínguez Cintas 2007, Lin et al. 2012, Bhattacharyya et al. 2015, Jett et al. 2020, Schöch et al. 2020). By removing the primarily copyright-relevant features from the original documents, it is possible to publish this derived data.

Until now, the derived data was expressed using formats and vocabularies based on each project's preferences. The transformation to derived data has considered mainly the textual information, giving less attention to other data (textual structure, metadata, or annotation). Many formats struggle modelling these kinds of data, which deteriorates their FAIR status (Wilkinson et al. 2016).

The German consortium Text+ is currently exploring derived data. We argue for the use of TEI as a format for derived data from texts, including different kinds of data into a single file. In contrast to other formats which are used only in specific

disciplines, TEI is known in many communities. While the conversion from TEI to these formats is possible, the inverse workflow is in many cases impossible. TEI allows fulfilling the FAIR criteria better than other formats, because it:

- offers elements to identify (F1) and describe (F2) the document,
- is an open format (A1.1), data and metadata are in a single file (A2),
- offers a rich vocabulary (I1) which follows FAIR criteria (I2), for the documentation of changes (I3),
- decisions (R1), licenses (R1.1), and origin of the data (R1.2).

To show our implementation, we use documents in five languages from already existing corpora:

- Gutenberg.de, the German clone of Project Gutenberg
- American drama on CD-ROM, a corpus of English plays (18th-20th centuries)
- The Spanish CoNSSA corpus (Calvo Tello 2021)
- A French corpus with 320 copyright protected novels
- A Chinese corpus with 158 texts by Lu Xun

In the talk, we will show several options of how researchers could model in TEI files transformed data from the original copyrighted texts. First, bag-of-words models can be expressed as measure elements and these can relate to different textual levels, such as volumes, chapters, paragraphs or even sentences. Second, the TEI elements within the body element could be understood as other kind of transformed data that can be used by researchers. The third option combines frequency of tokens and textual structure expressed through TEI elements by sorting the tokens randomly within a certain TEI element, such as paragraphs. However, this randomization spoils almost any calculation of n-grams or collocations, which are at the core of many current distributional NLP methods. Thus, our final solution is to create n-grams maintaining the original order of the tokens, and place them into a container (*seg*) which are randomly shuffled. These TEI files could be used with standard tools and still contain text (unreadable for humans), metadata, annotation, textual structure, and partly, the original distribution of the tokens.

## Bibliography

Bhattacharyya, Sayan, Peter Organisciak, and J. Stephen Downie. 2015. 'A Fragmentizing Interface to a Large Corpus of Digitized Text: (Post)Humanism and Non-Consumptive Reading via Features'. *Interdisciplinary Science Reviews* 40 (1): 61-77. <https://doi.org/10.1179/0308018814Z.000000000105>.

Calvo Tello, José. 2021. *The Novel in the Spanish Silver Age: A Digital Analysis of Genre Using Machine Learning*. Digital Humanities Research 4. Bielefeld: transcript. <https://www.transcript-verlag.de/978-3-8376-5925-2/the-novel-in-the-spanish-silver-age/?c=331025282>.

Lin, Yuri, Jean-Baptiste Michel, Erez Aiden Lieberman, Jon Orwant, Will Brockman, and Slav Petrov. 2012. 'Syntactic Annotations for the Google Books NGram Corpus'. In *Proceedings of the ACL 2012 System Demonstrations*, 169–74. Jeju Island, Korea: Association for Computational Linguistics. <https://aclanthology.org/P12-3029>.

Sánchez Sánchez, Mercedes, and Carlos Domínguez Cintas. 2007. 'El banco de datos de la RAE: CREA y CORDE'. *Per Abbat: boletín filológico de actualización académica y didáctica*, no. 2: 137–48.

Schöch, Christof, Frédéric Döhl, Achim Rettinger, Evelyn Gius, Peer Trilcke, Peter Leinen, Fotis Jannidis, Maria Hinzmann, and Jörg Röpke. 2020. 'Abgeleitete Textformate: Text und Data Mining mit urheberrechtlich geschützten Textbeständen'. Html,PDF,Xml. *Zeitschrift für digitale Geisteswissenschaften*. [https://doi.org/10.17175/2020\\_006](https://doi.org/10.17175/2020_006).

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. 'The FAIR Guiding Principles for Scientific Data Management and Stewardship'. *Scientific Data* 3 (March). <https://doi.org/10.1038/sdata.2016.18>.

## About the authors

**José Calvo Tello** works as a researcher and subject librarian at the Göttingen State and University Library. His research focuses on the application and development of computational and statistical methods to Romance literature and library records.

**Keli Du** is a researcher on the Zeta and Company and Text+ projects at the Trier Centre for Digital Humanities. His work focuses on computational literary studies. He is particularly interested in modelling and operationalising humanities research questions in data analysis problems.

**Mathias Göbel** is a Data Analyst at Göttingen State and University Library, one of the largest but definitely the greatest research library in Germany. Mathias

prepared the technical part of several TEI-based editions utilizing larger infrastructure provided by DARIAH-DE and TextGrid.

**Nanette Rißler-Pipka** is a digital humanist, literary scholar, specialist in French and Spanish literature. She is National Coordinator of Germany for DARIAH-ERIC. She works at the central office of the Max Weber Foundation-German Humanities Institutes Abroad.

# Five Aspects of ViFE

**Martin Albrecht-Hohmaier**, Paderborn University, Germany

**Irmlind Capelle**, Universität Paderborn, Germany

**Kristin Herold**, Paderborn University, Germany

**Johannes Kepper**, Beethovens Werkstatt, Germany

**Andreas Münzmay**, Paderborn University, Germany

**Salome Obert**, Carl Maria von Weber-Gesamtausgabe, Germany

**Kristina Richts-Matthaei**, Universität Paderborn, Germany

**Mark Saccomano**, Paderborn University, Center for Music, Edition, Media  
(ZenMEM)

**Agnes Seipelt**, Paderborn University, Germany

**Peter Stadler**, University of Paderborn, Germany

**Dennis Ried**, University of Paderborn

**Poster**

**Keywords:** Music Philology, Digital Humanities,  
Music Encoding, Text Encoding

## Abstract

The Virtual Research Group *Edirom* (*Virtueller Forschungsverbund Edirom*, ViFE) is an association of current and former staff of the Detmold/Paderborn Department of Musicology involved in projects advancing digital methods in musicology. With more than 25 researchers currently involved in around a dozen different projects, ViFE is a major player in musicological Digital Humanities.

## A Knowledge Sharing Group

The fundamental idea of ViFE is to be a group of experts in constant collaboration. Together, the expertise of ViFE members adds up to over 100 years of experience in digital music editions, ranging from computer science to philology and music history, from project management to Digital Humanities and information science – shared through weekly meetings and regular workshops within our group.

## Research Projects

ViFE is named after the *Edirom* project, funded from 2006 to 2012. ViFE is involved in long-term projects, including *Carl Maria von Weber Gesamtausgabe* and

*Beethovens Werkstatt*, but also has completed, or currently works on, several short-term projects and is constantly developing new project proposals.

## **Standards**

ViFE has been deeply involved in both the TEI and MEI communities for many years. We continue to actively connect our research with others and apply our findings to the development of these encoding standards.

## **Education and Training**

Since 2010, ViFE has organized the annual *Edirom Summer School*, one of the most important venues for learning MEI. Members of ViFE are also involved in Master's programs on Digital Music Philology and on Digital Humanities at Paderborn University and supervise doctoral dissertations.

## **Cooperation**

Our group is a vital part of NFDI4Culture and partners with many other institutions, such as Mainz Academy of Science and Literature, cemfi, CDMD, RISM, Beethoven-Haus Bonn, SLUB Dresden, Staatsbibliothek zu Berlin, BSB Munich, University of Oxford e-Research Centre, Gesellschaft für Musikforschung, and several German universities.



# From TEI to CIDOC-CRM with LINCS XTriples: Preserving Meaning across formats?

Constance Crompton, University of Ottawa

Alice Defoure, University of Ottawa, Canada

Short Paper

**Keywords:** linked data, tool development, CIDOC-CRM, infrastructure

## Abstract

This demonstration paper will introduce the Linked Infrastructure for Networked Cultural Scholarship (LINCS) project's beta XTriples tool. LINCS XTriples converts TEI to CIDOC-CRM. The original XTriples, developed by Torsten Schrader from the Academy of Sciences and Literature, Mainz, was designed to convert XML to RDF (both ttl and XML), provided that users could create configuration files that describe the mapping between their input files and desired output. A complement to more sophisticated mapping tools, such as FORTH's 3M Editor, our version of XTriples has been redesigned for TEI users who are unlikely to have a deep knowledge of CIDOC-CRM but who would like to contribute their data to LINCS or who would like CIDOC-CRM for their own projects. We have stripped out the mapping functionality, but offer a series of XSLTs that do the mapping work, pulling in URIs from TEI created in LINCS' LEAF-Writer to produce CIDOC-CRM that conforms to LINCS' application profiles (Jakacki and Brown).

LINCS XTriples follows on a long tradition of work to create RDF from TEI (Eide et al., Tomasi et al., Šimek). The challenge in this work comes from RDF's flatness relative to TEI-encoded documents. The meaning of TEI-encoded text comes from the hierarchical nature of the language. Any one element in TEI may be drawing its meaning from its relationship to a grandparent well up the XML tree. RDF, by comparison is fairly flat, and when working with a reasonably tight ontology, it offers usefully narrow sets of meaning for each element and attribute value. In addition to our demonstration of the conversion of TEI to CIDOC-CRM, we will discuss the challenge of moving from TEI's relatively simple ISO encoding of time to CIDOC-CRM's rich, nuanced, and URI-dependant approach to time.

## Bibliography

Eide, Øyvind, et al. *Encoding Cultural Heritage Information for the Semantic Web. Procedures for Data Integration through CIDOC-CRM Mapping*. 2008 pp. 47–53.

Jakacki, Diane, and Susan Brown. *LEAF-Writer: Low-Barrier Text Encoding for Teaching and Editing*. MLA, 2023.

<https://mla.confex.com/mla/2023/meetingapp.cgi/Session/15104>.

Šimek, Jakub. *Referencing an Editorial Ontology from the TEI. An Attempt to Overcome Informal Typologies*. <https://doi.org/10.5281/zenodo.3457974>.

Tomasi, Francesca, et al. *Dialogue and Linking between TEI and Other Semantic Models*. TEI Conference 2013.

# Half-day tutorial: Music Encoding with mei-friend

**David M. Weigl**, mdw – University of Music and Performing Arts Vienna, Austria

**Werner Goebel**, mdw – University of Music and Performing Arts Vienna, Austria

**Anna Plaksin**, Johannes Gutenberg-University Mainz, Germany

**Stefan Münnich**, University of Basel, Switzerland

**Mark Saccomano**, Paderborn University, Center for Music, Edition, Media  
(ZenMEM)

**Workshop**

**Keywords:** MEI, mei-friend, introduction,  
tutorial

## Abstract

### Topic

The Music Encoding Initiative's MEI XML schema supports the comprehensive, machine-readable, semantic capture of a large range of music documents in a variety of notation formats. However, the learning curve may be steep, particularly for users lacking prior expertise with XML – and an absence of prior exposure to such technologies is not unusual among the audience most likely to benefit from music encodings, that is, music librarians, scholars, and performers. Unlike text, music suffers from the additional complication that its encodings necessarily abstract far away from the source; while a text-encoded paragraph (e.g., a TEI `<p>` element) contains recognisable textual content, the `<note>` elements within an MEI `<measure>` look nothing like their notated equivalents in music scores.

Typical workflows incorporating MEI have involved frequent iterations of: XML editing (in an editor such as Oxygen); running the MEI through the Verovio engraving tool to visually inspect the encoded notation; and returning to the XML editor for fixing, refinement, and continued encoding. The mei-friend Web application, available via <https://mei-friend.mdw.ac.at>, is a tool developed to simplify this traditional workflow. It exposes two tightly-coupled panels that respectively incorporate an XML editor (provided by CodeMirror) and a digital music score (rendered by Verovio), supporting synchronised interaction, navigation, and editing in either modality. It also integrates important infrastructural conveniences,

including schema-based MEI validation and autocompletion; Git cloud-service integration with GitHub and Git Lab for collaborative encoding; Solid (RDF) integration for distributed score annotation; a built-in MIDI player, for sonification and auditory validation of the encoding; and, a direct interface to the MEI Guidelines to look up documentation for selected elements. A dedicated 'Help' website offers extensive documentation of mei-friend's basic and advanced features. The application aims to lower the barriers to entry to the world of music encoding, while providing a wealth of features and optimisations to facilitate and speed up the encoding process for experienced music editors.

In this tutorial, we will provide a gentle introduction to encoding music with mei-friend, walking participants through the task of encoding a simple melody from scratch using MEI, before opening up to a broader consideration of 'real-world' music encoding workflows. Tuition will be based around 'hands-on' exercises, with experienced co-conveners available to assist participants when help is required.

## **Audience**

This tutorial primarily caters to participants who are new to the world of music encoding, and will be well-suited both to experienced users of TEI, as well as to those new to XML-based encodings. In addition, the tutorial will be of interest to experienced users of MEI who are curious about the affordances of mei-friend.

## **Requirements**

### **Participants**

Participants will be required to bring their own laptops (Windows / Linux / Mac), optionally with headphones. No software installation will be required.

### **Room set-up**

We will require a projector and A/V set-up that permits plugging in a presenter laptop (HDMI or USB-C). Participants will require tables and some shared power sockets for their laptops, and seating should be arranged to allow tutorial conveners access to assist individual participants during the hands-on sessions. All participants should be able to view a presentation screen, to allow them to follow along.

## Schedule and topics

The tutorial's content will fit a half-day (3 hour) session scheduled as follows:

15 min: Introduction to music encoding formats and workflows 1 h 15 min: *Hands-on*: Guided walk-through of a simple music encoding task

20 min: Break

15 min: Whistlestop tour of mei-friend and its documentation 10 min: Real-world encoding workflows (OMR, MusicXML conversion, etc) 25 min: *Hands-on*: Real-world encoding task – Cleaning up Beethoven's Op. 76 20 min: Questions and open discussion

## About the authors

**David M. Weigl** is a postdoctoral research associate at the IWK – mdw with a focus on applications of semantic Web technologies in digital music research. He co-develops the mei-friend Web application, and has been involved in teaching MEI using mei-friend in conference tutorial and class-room contexts.

**Werner Goebel** is a music researcher with background in musicology, psychology, and computer science, and head of the Department of Music Acoustics – Wiener Klangstil (IWK) at mdw – University of Music and Performing Arts Vienna. He originated mei-friend in its first incarnation as an Atom plug-in, and continues its development as a Web application.

**Anna Plaksin** is a postdoctoral research associate at the Institute of Art History and Musicology (IKM) at Johannes Gutenberg University Mainz. She currently works on support for editorial markup in the mei-friend Web application.

**Stefan Münnich** is a research associate at the Anton Webern Gesamtausgabe, Basel, where he explores ways to apply RDF-based semantic models for the purpose of a scholarly digital music edition. He was recently appointed technical co-chair of the MEI and co-developed the MEI tutorials.

**Mark Saccomano** is a music theorist at Paderborn University. He previously taught music history and music theory at Columbia University and was adjunct professor of music at Montclair University in New Jersey. He is currently working as a digital musicologist on the Beethoven in the House project.

# “Herzlichst in Eile Ihr Mahler” – Indexing the letters of Gustav Mahler

Clemens Gubsch, Austrian Academy of Sciences (ÖAW), Austria

Raphaela Quass, ÖAW, Austria

Poster

**Keywords:** correspondence, indexing, metadata

## Abstract

Compared to other composers' extant letter exchanges, Gustav Mahler's correspondence with approximately 5000 letters is one of the most extensive letter collections of the 19<sup>th</sup> and 20<sup>th</sup> centuries. In addition to family mail, written exchanges with friends, artists, publishers, artistic directors and institutions, the correspondence also includes intellectual dialogues with other composers such as Richard Strauss or conductors such as Felix Weingartner. The resulting network of letter correspondents locates, on the one hand, Mahler's influence in the historical context while simultaneously making his sociocultural position as well as the accompanying period-specific variance recognizable. On the other hand, the number and content of the letters characterize Mahler as a person in a synchrisis between the composer's external image and the perception of his contemporaries.

The project therefore aims to create a basis for an edition of letters by recording and indexing all of Gustav Mahler's letters that have so far appeared in print as well as the existing counter-letters, which means only metadata of the letters and bibliographic records of the printed editions are recorded but no full texts. The encoding scheme of the TEI is used for the recording. Particular difficulties lie, on the one hand, in the harmonisation of the data collected within the various letter editions, i.e. classification of the letter formats (postcards, telegrams, contracts, etc.) or the reference to the place where the letters are kept (persistent library information) and, on the other hand, in Mahler's habit of not dating letters or, in some cases, dating them incorrectly as well as the small stock of counter letters. The poster will present the results of the project as well as the difficulties that have arisen in the recording, evaluation and presentation of the index of letters.

# Hybrid music editions – challenges at the Reger-Werkausgabe in retrospective.

Nikolaos Beer, University of Paderborn; Max-Reger-Institut/Elsa-Reger-Stiftung, Karlsruhe

Long Paper

**Keywords:** Digital music edition, workflows, project definitions, challenges

## Abstract

After almost 16 years of hybrid scholarly music edition work at the Reger-Werkausgabe<sup>[1]</sup> (RWA) at the Max-Reger-Institute<sup>[2]</sup> (MRI) and two years ahead of the project's completion, this paper provides retrospective insights concerning the initial ideas about the type, scope, and content dependency of the edition's digital part to the rather traditional edition work that materializes itself in its printed volume(s). It will outline that the situation today is at most different from the situation at the beginning where the printed part was seen as a "parent" to its new little "digital" companion.

The RWA (2008–2025) – funded by the German "Union of the Academies"<sup>[3]</sup> within the "Academies Program"<sup>[4]</sup> – was one of the first long-term scholarly music edition projects to merge and essentially integrate digital research and publication workflows based on MEI and TEI data facilitated in the Edirom<sup>[5]</sup> Tools, with a predominantly print-based project definition. The project has since evolved from a community-counseled "user" within^ the digital turn to a "self-supporter" and "self-producer" of digital workflows, data sets and publications. ^

Initially, the project's focus on its digital part was lying on questions about digital or digitizing transformation, presentation, and contextualization of print-based edition work. After a thorough phase of defining workflows and software tools, a similarly streamlined and economized publication process should have been emerged as was already the case with printed editions for decades. But trying to simply digitize working and publishing conventions from printed editions proved to be a fallacy. The main reasons were:

- *Content creation and ways of digital research dissemination:* The change of medium not only opened an infinite space for simply more content but exposed some rather media-dependent than content-related conventions of addressing and (over-)compressed presentation of music edition problems.
- *Technological progress:* Technological decisions made at one point in the project's development had to be continuously retuned and reevaluated, often causing inevitable delays.
- *Embedding and long-term preservation/availability of data and publications:* Being located at the MRI, the RWA not only reuses and works on decades of musicology research stored within the institute's archives, collections, and library. In fact, it leads to challenging digital feedbacks within the MRI's infrastructure and its near and intermediate future.
- *Human resources:* Since its beginning the project has managed to extend its digital human resources from a student assistant to a full-time "digital edition research engineer". But mastering the interplay of the vast amount of research content and its digital presentation, communication and preservation remains one of the project's main challenges to date.

The paper will not only name and describe the challenges but try to filter and propose some "lessons learned" concerning possible consecutive projects. The author has been responsible part of the RWA and responsible for its digital parts throughout its entire run.

## Notes

---

1. Reger-Werkausgabe, project page and publication portal „RWA Online“: <https://www.reger-werkausgabe.de>; last seen 30.4.2023. ↩
2. Max-Reger-Institut/Elsa-Reger-Stiftung, <https://www.max-reger-institut.de>; last seen 30.4.2023. ↩
3. Union der Deutschen Akademien der Wissenschaften, <https://www.akademienunion.de/>; last seen 30.4.2023. ↩
4. Akademienprogramm der Union der Deutschen Akademien der Wissenschaften,



<https://www.akademienunion.de/forschung/akademienprogramm>; last seen 30.4.2023. ↩

5. Edirom – Digitale Musikedition, project description, <https://edirom.de/edirom-projekt/>; last seen 30.4.2023. ↩

# Integrating TEI and MEI

**Agnes Seipelt**, Paderborn University, Germany

**Kristin Herold**, Paderborn University, Germany

**Ran Mo**, Paderborn University, Germany

**Dennis Friedl**, Paderborn University, Germany

**Workshop**

**Keywords:** MEI, TEI, ODD

## Abstract

It is a common situation to find music included in otherwise textual documents. These inclusions may range from individual musical symbols like single notes to full scores, bound as separate pages into a larger book. While these extreme cases are relatively straightforward to handle in an encoding, thanks to the availability of music fonts based on the SMuFL standard on one end, and the use of different files for the other, the middle ground seems much harder: a small melody sketched into a letter, a four-measure example printed into a book on music, etc. While the TEI offers the `<tei:notatedMusic>` element, by default it treats such music inclusions like images, providing a caption and links to images. In addition, the TEI Guidelines for this element also contain a note concluding with the remark that “it is also recommended, when useful, to embed XML-based music notation formats, such as the Music Encoding Initiative format as content of `notatedMusic`. This must be done by means of customization.” However, no example is given for such a customization, leaving the user alone with this task.

In this workshop, we will introduce examples of documents with mixed music/text content. We will encode these examples using a combination of MEI and TEI and will learn about how to validate this with a custom schema based on an ODD customization integrating both standards. Finally, we set up a small demo website that will render both TEI and MEI to HTML. Participants are expected to have some familiarity with at least one of TEI or MEI; the examples used will be simple enough to understand the encoding model of the other standard at sight. Some familiarity with the idea of ODD is certainly helpful, but intimate knowledge of the technical details is not necessary.

This will be full day hands-on workshop, and the participants are expected to bring their own device, running a validating XML editor. In the workshop, XSLT

stylesheets will be applied to XML documents. Instructors will use the Oxygen XML editor for this, which has a free evaluation period. Assistance on other products cannot be guaranteed.

## About the authors

**Dennis Friedl** ([dennis.friedl@uni-paderborn.de](mailto:dennis.friedl@uni-paderborn.de)) studied history and digital humanities at Paderborn University (UPB), where he specialised in digital scholarly editions. He now works in research data management at UPB and is responsible for the digital edition of the *Erich Wolfgang Korngold Werkausgabe*. He is a local organizer of the conference.

**Kristin Herold** ([herold@beethovens-werkstatt.de](mailto:herold@beethovens-werkstatt.de)) studied musicology with minors in philosophy and economics at the TU Berlin. Research assistant at the Musikwissenschaftliches Seminar Detmold/Paderborn 2011-2016 in the *DARIAH-DE* project, 2012 freelancer for the digital *Carl-Maria-von-Weber-Gesamtausgabe* and 2016 research assistant in the *Hoftheater Project*. 2013-2016 Research assistant in the *Sarti project* at the UdK Berlin. 2016-2018 research assistant at the *B. A. Zimmermann-Gesamtausgabe*. Research assistant at the Musikwissenschaftliches Seminar Detmold/Paderborn 2018-2020 in the project *Pasticcio. Ways of arranging attractive operas* and since 2018 in the project *Beethovens Werkstatt*.

**Ran Mo** ([mo@beethovens-werkstatt.de](mailto:mo@beethovens-werkstatt.de)) studied musicology at the Musikwissenschaftliches Seminar Detmold/Paderborn and the Paderborn University. She completed the master's program in 2017 with a thesis on materiality issues in music editing using the example of engraving tools from the Schlesinger publishing house and developed methods and tools for dating and interpretation questions. From 2017 to 2019 she worked as a research assistant in the projects *Beethovens Werkstatt*, *Carl Maria von Weber Gesamtausgabe*, and *Detmolder Hoftheater*. Since 2019 she works at *Beethovens Werkstatt*.

**Agnes Seipelt** ([seipelt@beethovens-werkstatt.de](mailto:seipelt@beethovens-werkstatt.de)) studied Musicology at the Musikwissenschaftliches Seminar Detmold/Paderborn and the Paderborn University. She completed the master's program in 2017. From 2017 to 2019 she worked as a research assistant in the project *Digital music analysis with MEI using the example of composition studies by Anton Bruckner* in cooperation with the Austrian academy of sciences. Since 2019 she works as a research assistant at *Beethovens Werkstatt*.

# Interactive music encoding with Verovio

Laurent Pugin, RISM Digital Center

Andrew Hankinson, RISM Digital Center, Switzerland

**Workshop**

**Keywords:** Music encoding, interactive applications, MEI, Python, JavaScript

## Abstract

Verovio (<https://verovio.org>) is an open-source software library for engraving music scores encoded in MEI. It is a flexible tool that can be used in a wide range of applications. Projects use Verovio in the context of digital critical editions, genetic editions, early music, audio alignment, online editing, music addressability, visualization, performance, and even composition via artificial intelligence. Verovio is written in C++, but it can also be used in diverse programming ecosystems, including as a JavaScript toolkit in web-pages, or as a module in Python environments.

One of the unique features of Verovio is that it renders music notation into SVG where the original structure of the MEI is preserved as much as possible. Not only is the original XML tree structure of the MEI preserved in the SVG, but also element XML IDs and element names through corresponding SVG class names. Verovio features a set of methods for retrieving information about MEI elements in the rendering context, including time-related information in MIDI sound files—another output format that Verovio can produce. This design allows the SVG music notation to serve as an interaction layer for building dynamic applications of various kinds.

In this full-day workshop, participants will be provided with an overview of the Verovio project and its philosophy. The overview will cover different use-cases and explain how Verovio can be integrated in an application. You will also learn about the current development status and about the latest additions. Most importantly, you will follow a step-by-step introduction on how to integrate Verovio in a JavaScript web environment, from the most basic use case to more advanced, interactive, and project specific ones. A similar approach will be provided for Python integration, with some examples on how music encoding data can be processed in scripting environments.

There are no specific prerequisites for attending the workshop. However, we believe it would be more interesting for people having at least one of the following: some understanding of MEI (or another music encoding scheme), some knowledge of music notation, or some minimal programming skills (JavaScript or Python). Participants are expected to bring their own laptop. No specific tool is required since we will be using online tools.

The learning outcomes for the workshop will be:

- to understand the overall architecture of Verovio and the ways it can be used.
- to know how to create a minimal webpage integrating Verovio rendering.
- to understand how the SVG output in Verovio acts as an interaction layer.
- to know how to interact with the music notation in the webpage with JavaScript and CSS.
- to understand how to load Verovio in Python scripting environments.
- to know how to write a minimal script that processes music encoding files with Verovio.

## About the authors

**Laurent Pugin** is co-director of the RISM Digital Center and is the lead developer of Verovio. Previously, he was a postdoctoral researcher at McGill University and a visiting scholar at Stanford University. He is also co-applicant of LinkedMusic, a research project based at McGill University. Laurent Pugin has been a member of the MEI board and co-chair of its technical team since 2014. In 2019, he completed a habilitation in both musicology and digital humanities at the University of Bern where he has been teaching for many years.

**Andrew Hankinson** is senior software developer at the RISM Digital Center where he leads the development of RISM Online. He was a postdoctoral researcher at the University of Oxford and senior software engineer for Digital Collections Discovery at the Bodleian Libraries. Andrew specializes in web technologies for building scalable and usable search and retrieval systems. He has held positions on the MEI board and with the Executive of the IIF Consortium. He has been an active contributor to MEI for many years.

# Interoperability of Text and Image in the Digital Edition of the Wenceslas Bible

Linda Beutel-Thurow, University of Salzburg

Julia Hintersteiner, University of Salzburg

Manfred Kern, University of Salzburg

Oleksii Sapov-Erlinger, Mozarteum Foundation Salzburg, Austria

Poster

**Keywords:** digital edition, cultural history, text-image-correlation, TEI, ICONCLASS

## Abstract

The *Wenceslas Bible*, created for the Bohemian King Wenceslas IV. around 1390-1400, is the first largely complete German-language translation of the Hebrew Bible after the Latin Vulgate, combined with elaborate and highly valuable illustrations. Therefore, it is one of the most precious cultural heritage objects of the Austrian National Library. The Bible contains six codices (Cod. 2759-2764) with 1214 parchment leaves featuring 654 main and countless marginal miniatures. The text of the Bible translation is of the highest interest for philologist and eminently significant in terms of cultural history, as it is integrated into the religious reform movements of the 14th century and must be seen in connection with the so-called Austrian Bible translator, the work of Jan Hus but also the English Wycliffite translation.

*The Wenceslas Bible – Digital Edition and Analysis* is a joint project of the *University of Salzburg* and the *Austrian National Library*, financed by the Federal State of Salzburg. It aims to create a TEI based edition with a focus on the investigation of the text-image correlations. As a first step, the project has launched a [prototype](#) for the book Genesis in June offering a synoptic view for transcription and the facsimile. Further parts will be published continuously, including a revised text with editorial comments as well as a description and analysis of the illustrations including their relation to the text.

The poster will delineate the primary goals of the project as well as the significant difficulties related to the process and compatibility of text modification and digital

picture encoding. A section of the display will exhibit the TEI model implemented to connect the two edition tiers and the necessary connections that must be established throughout. It underscores the cross-disciplinary nature of the edition, encompassing methodologies from philology, art history, and Digital Humanities, with a particular emphasis on the methodical correlation between text and image, and its digital portrayal.

## About the authors

**Linda Beutel-Thurow** has been a research assistant at the University of Salzburg in the project [The Wenceslas Bible – Digital Edition and Analysis](#) since 2021. She is also writing her dissertation in the field of older German language and literature on the topic of Naming poetics of Middle High German epic. In both areas, she works thematically in the field of the semantic web and specializes in the use of tools such as Transkribus and AntConc, the TEI data format and norm data such as Wikidata.

**Julia Hintersteiner** is a Digital Humanities major and has been working as a research assistant at the [The Wenceslas Bible – Digital Edition and Analysis](#) since September 2022. She is working on image annotation and digital image recognition in the project, especially focusing on text-image correlations. Her interests include digital art history and intermediality in medieval context.

**Manfred Kern** is a full professor for medieval German language and literature at the University of Salzburg since 2010. His research concentrates on the reception of antiquity in the medieval ages, the theme of vanity in medieval literature, comparative studies in medieval lyric and epic poetry and recently in the medieval German bible translation. He is head of the project [The Wenceslas Bible – Digital Edition and Analysis](#). For further information: [www.plus.ac.at/germanistik/der-fachbereich/mitarbeiterinnen/kern-manfred/](http://www.plus.ac.at/germanistik/der-fachbereich/mitarbeiterinnen/kern-manfred/)

**Oleksii Sapov-Erlinger**, MA, ([0009-0003-7012-5781](tel:0009-0003-7012-5781)), has been working since 2018 as a research assistant at the [Digital Interactive Mozart Edition](#), a joint project of the *Mozarteum Foundation Salzburg* and the *Packard Humanities Institute*, with focus on Data Modeling and Data Engineering in MEI. Additionally, since July 2022, he has been working as a research assistant at the [Die Wenzelsbibel – Digitale Edition und Analyse](#), a joint project of the University of Salzburg and the Austrian National Library. His focus in this position is on TEI Data Engineering and Web Development.

# Introduction to publishing XML with static site and front-end technologies

Raffaele Vigiante, University of Maryland

Workshop

**Keywords:** digital publication, minimal computing, JavaScript, infrastructure

## Abstract

### Scope of the workshop

This half-day workshop will introduce strategies for handling XML when publishing with static site generators and front-end technologies. The workshop will focus on isomorphic approaches to publishing XML data on the web or, in other words, publishing XML data with little or no transformation, or with a structure-preserving mapping that allows working with the output as if it were the initial data source.

In particular, attendees will be introduced to CETELcean as a way of publishing TEI with minimal (or no) transformation (Cayless and Vigiante 2018), and Verovio, a rendering engine for MEI that produces isomorphic SVG output (Pugin 2016). The main focus of the workshop will be learning how this approach can be used in conjunction with static site generators and will work on examples in “vanilla” JavaScript, React, and Gatsby. This workshop is aimed at attendees who already have some experience with programming (including XSLT) and the command line; however, all are welcome and will be supported as much as possible throughout the workshop.

### Motivation

Digital humanities projects that result in the creation of digital output—typically a website—are prone to what Smithies et al. call the “digital entropy of software and digital infrastructure” (2019). Static sites have become a common choice for archiving legacy projects that risk going offline (Smithies *et al.* 2019, Summers 2016) because they only require the absolute minimum from hosting infrastructure: a server to distribute documents at a given address. The sites themselves, once created, require no active maintenance and can be easily moved and transferred



like any other collection of files. However, static sites cannot support features that would require an active server, such as large scale text search and user management; these features, therefore, are removed when projects are archived into static sites. Deriving static sites from an end-of-life project is the clear choice when access to infrastructure becomes limited. But this workshop addresses the question: What would it take to adopt static sites from the start?

## Schedule and requirements

After a brief introduction on static sites and the motivations for using them, the workshop will introduce:

- Exploring isomorphic approaches: CETELcean and Verovio in “vanilla” JavaScript.
- How to extend CETELcean with React.
  - Optional collaborative work: experiment with other frameworks that attendees may be familiar with (e.g. Vue, Angular, etc).
- Optionally, if attendees are interested in music notation: using Verovio in React for interactivity.
- Using the static site generator Gatsby and its TEI plugins.
- Keeping within a single stack: applying optional transformations to TEI documents in NodeJS (using JavaScript / TypeScript or XSLT via SaxonJS).

Example TEI and MEI documents to integrate into the website will be provided, but attendees are encouraged to bring their own.

Participants must bring their own laptop and be able to install (free) software on it. Internet access will be required. The tutor will require a projector.

## Bibliography

Cayless, Hugh, and Raffaele Vigiante. 2018. “CETELcean: TEI in the Browser.” Presented at Balisage: The Markup Conference 2018, Washington, DC, July 31 – August 3, 2018. In: *Proceedings of Balisage: The Markup Conference 2018. Balisage Series on Markup Technologies*, vol. 21.  
<https://doi.org/10.4242/BalisageVol21.Cayless01>.

Pugin, Laurent. 2016. "Interaction with Music Encoding". In: „*Ei, dem alten Herrn zoll' ich Achtung gern*“ *Festschrift für Joachim Veit zum 60. Geburtstag*. Eds. Kristina Richts and Peter Stadler. 617-631. Allitera Verlag.

Smithies, James, Carina Westling, Anna-Maria Sichani, Pam Mellen, and Arianna Ciula. 2019. "Managing 100 Digital Humanities Projects: Digital Scholarship & Archiving in King's Digital Lab." In: *Digital Humanities Quarterly* 013 (1).

Summers, Edward. 2016. "The Web's Past Is Not Evenly Distributed." Maryland Institute for Technology in the Humanities (blog). May 27, 2016.  
<https://mith.umd.edu/webs-past-not-evenly-distributed>.

## About the author

**Dr. Raffaele (Raff) Viglianti** is a Senior Research Software Developer at the Maryland Institute for Technology in the Humanities, University of Maryland. His research is grounded in digital humanities and textual scholarship, where "text" includes musical notation. He researches new and efficient practices to model and publish textual sources as innovative and sustainable digital scholarly resources. Dr. Viglianti is currently an elected member of the Text Encoding Initiative technical council and the Technical Editor of the Scholarly Editing journal.

# Jazz music encoding for education: the collaboration between MEI and TEI.

Patricia García-Iasci, Universidad de Alicante, Spain; Universidad de Salamanca, Spain

Javier Merchán Sánchez-Jara, Universidad de Salamanca, Spain

Short Paper

**Keywords:** Representation of musical message, MEI, TEI, Jazz encoding, American cipher

## Abstract

The representation of the musical message has varied according to the period, the media, the musical genre and the method of composition or at least the context of notation. These systems of musical representation are based on a code composed of several symbols, letters and numbers that represent something, in this case it has a musical meaning. In order to understand the relationship between musical meaning and code, an interpretation is necessary and this interpretation must have its rules.

In Jazz, the American, Anglo-Saxon, English or modern cipher is used as the reduction to the minimum expression with meaning; a "musical common minimum". Roughly speaking, it consists of capital letters, which are used to represent the root note. The absence or not of certain numbers accompanying this letter provides information about the type of chord, and symbols such as -, +, b, or # and dimensions such as aug, min, dim, give more information about the types of chords (Peñalver Vilar. J.M, 2010).

Each cipher is oriented to certain objectives: the American cipher provides the minimum musical information for a performance based on improvisation. The functional cipher provides harmonic information for compositions and chord progressions and tonal centres for compositional purposes. Both ciphers are represented by alphanumeric characters and have a musical meaning without the need to be located on staves, as is the case with notes.

By combining TEI and MEI, we could make a multi-purpose musical representation. With TEI, the more "literary" part of the American and functional cipher can be

encoded. With MEI, we can make a complete musical representation of the harmonic notes of Jazz on the staff with melody and rhythm. This results in a representation composed of a Standards representation of the RealBook, FakeBooks, or lead-sheet and annotations for interpretation and functional analysis annotations that can be very useful for questions of adaptations in interpretation or beneficial information for improvisation and composition (Young. H, 2017).

In order to carry out this research, use has been made of selected works from the University of Alicante to train an OMR tool. These are selected Jazz Standards by authors such as Django Reinhardt, Ella Fitzgerald, George Gershwin, Fats Waller, Duke Ellington, Thelonious Monk, Dizzy Gillespie, John Coltrane, Miles Davis, etc... There are a total of 226 examples that we hope will be enough to create an effective sample.

The purpose of this idea is a larger project that aims firstly to recognize handwritten jazz standards by means of the OMR, and secondly to extract semantic information about how these chords work. (Calvo-Zaragoza et al. 2020) Thirdly, to use this information to create music, practice it and study it for educational purposes. This paper focuses on the second part mentioned above, on the search for the extraction of this semantic information and its representation (Harte et al. 2005).

In this way, together with functional analysis, we would approach the possibility of creating a grammar for multiple applications, such as harmonic progressions, melody creation and accompaniment (Steedmann, 1984). These applications would then be placed in the educational field, in order to integrate coding and digital media into the pedagogical practices of musicians and musicologists (Duguid, 2020).

## Bibliography

Calvo-Zaragoza, J., Jr, J. H., and Pacha, A. (2020). 'Understanding optical music recognition'. *ACM Computing Surveys (CSUR)*, 53(4), 1-35.

Duguid, T. (2020). 'The forgotten Classroom? Bringing Music Encoding to a New Generation'. In: *Music Encoding Conference 2020*, 26-29 May 2020.

Harte, C. C, Sandler, M. B, Addallah, S. A, & Gómez, E (2005, September). 'Symbolic Representation of Musical Chords: A proposed Syntax for Text Annotations'. In: *ISMIR* (Vol. 5, pp. 66-71).

Peñalver Vilar, J. M. (2010). *El cifrado armónico en el jazz y la música moderna: un recurso imprescindible para la notación musical*. Universitat Jaume I de Castelló.

Steedman, M. J. (1984). 'A generative grammar for jazz chord sequences'. In: *Music Perception*, 2(1), 52-77.

Young, H. (2017). 'A categorial grammar for music and its use in automatic melody generation'. In: *Proceedings of the 5th ACM SIGPLAN International Workshop on Functional Art, Music, Modeling, and Design*.

## About the authors

**Patricia García-Iasci** is a PhD student at the Salamanca University. She holds a master degree in Digital Humanities, focusing specifically on music. She is working in PolifonIA and researches and collaborates in projects related to optical music recognition and music encoding with the University of Alicante and University of Salamanca.

**Javier Merchán Sánchez-Jara** is a Doctor Professor in the Department of Didactics of Musical, Plastic and Corporal Expression and is a Responsible Researcher GIR DIDEROT. And Coordinator of the Music specialty of the MUPES Master at the University of Salamanca.

# Lessons from the Classroom: MEI for Data Scientists

Richard Freedman, Haverford College, United States of America

Daniel Russo-Batterham, University of Melbourne, Australia

Long Paper

**Keywords:** MEI, XML, Data Analysis, Pedagogy

## Abstract

Many Data Science and Computer Science students today are familiar with JSON, and may even have worked with APIs to extract data from the web. Ask about XML, however, let alone TEI or MEI, and you are often met with quizzical looks. Yet XML files contain much information that can be productively analyzed with modern Data Science tools, so training students to leverage these materials is a worthwhile endeavor. Our work on *Citations: The Renaissance Imitation Mass* (CRIM; [crimproject.org](http://crimproject.org)) has demonstrated the value of using MEI as the basis for representing a digital score (with the help of Verovio) *and* performing sophisticated queries over large corpora. Encodings can thus be created with these dual uses in mind.

In this presentation, we reflect on our experience teaching XML concepts in the Data Science classroom with—for largely practical reasons—Python tools already familiar to students, and talk through an indicative analysis. A novel approach we used to explain the hierarchical structure of XML was through colorful and interactive network visualizations. This helped students to quickly grasp not just the tree-like structure of XML, but the scale and complexity of an encoded text, as well as the diversity of the elements used and where they appear, with some elements frequently “leaf” nodes and others structuring much broader sections.

Once armed with this understanding, they could begin to analyze the attributes contained within these elements in ways that reveal new insights, from the distributions of features, to patterns of similarity; with one caveat: all their wrangling, analysis, machine learning and visualization tools expect data in a tabular format.

Students were then asked to extract select information into tables with BeautifulSoup, while thinking about the implications of representing the same knowledge in different ways, in this case through varied data structures. What was lost in translation when going from one structure to another? How did the way the XML was originally created affect this process and the final results of analysis? How does hand-coded XML differ from machine-generated? What would they do differently when creating their own XML if they knew it would be used for analysis? What, if anything, can we learn from these questions about the varied means of encoding knowledge and culture more broadly, from text (ink), to music (sound), to art (paint)?

In answering these questions, we will retrace some of the key steps of our pedagogy, moving between macro- and micro-level views of MEI documents and the encoding practices they embody. We will begin by viewing the hierarchy of an XML document as a network of nodes and edges—a representation that gives students a vivid understanding of both the complexity and the logical structure that otherwise appears to them as just another text document in an editor. From here we will zoom in on some key nodes in the network, showing how students began to traverse the XML files intelligently in order to answer various kinds of queries about these texts, the musical concepts they encode, and how the structured XML encoding compares with scores in staff notation.

One productive point of inquiry centers on accidentals—chromatic inflections that happen from time in European music as pieces move between different diatonic pitch spaces. In Early Music in particular, the original sources were often less than explicit. Composers, scribes, and printers often assumed that competent singers and instrumentalists would simply know to add a sharp, flat, or natural before some given note depending on its melodic and contrapuntal context. In modern transcriptions these unwritten accidentals (typically called *musica ficta*) are often made explicit. Graphical editions do so by putting these symbols above the staff, or by surrounding them with parenthesis. But in the case of MEI encodings we can be still more precise, since `<supplied>` elements can convey all sorts of detail about the meaning, intention, and ultimate responsibility for such work. Comparing the notated and supplied accidentals in a given piece can tell us something about the explicitness of that particular edition; doing so across a corpus can reveal biases of editorial practice. And these kinds of lessons in turn, prompted interesting reflections on the part of students about the biases and values implicit in all editorial work, and the technologies that represent the perishable art of music.

All materials will be available as public Jupyter notebooks for reuse in classrooms and laboratories.



# Lights, Camera, Slugline? Encoding Screenplays in TEI

Joey Takeda, Simon Fraser University, Canada

Sydney Lines, University of British Columbia, Canada

Mary Chapman, University of British Columbia, Canada

Long Paper

**Keywords:** screenplays, film, performance, drama

## Abstract

The Winnifred Eaton Archive is an accessible, fully searchable, digital scholarly edition of the collected works of Winnifred Eaton Babcock Reeve, an early Chinese Canadian author now best known for the popular Japanese romances she signed “Onoto Watanna” in the early twentieth century. She is less frequently recognized, however, as an important contributor to Hollywood’s “Golden Age,” during which she served as a scenarist and editor for Universal and MGM Studios (Chapman and Lines, 2023). Though Eaton’s involvement with the film industry has received some acknowledgment (such as her role in Universal’s *Undertow* (1930) and *Young Desire* (1930), both starring Mary Nolan), the WEA has recently recovered numerous screenplays, scenarios, adaptations, and other cinematic texts written by Eaton in the 1920s and 1930s, which the project has been working to encode, alongside the rest of Eaton’s corpus, in TEI P5 (4.6.0).

Screenplays are complex literary, textual, and bibliographic artifacts that follow strict structural rules and formatting conventions, which began to emerge in Eaton’s era and became formalized by the 1950s through the Writers Guild of America (Price 2013; Johnson 2019). However, while the TEI’s “drama” module (and described in “Performance Texts”) offer a handful of elements for encoding “Other Types of Performance Text” (7.3), there is little sustained attention to the forms and features unique to screenplays and their antecedents (such as scenarios and treatments), and, as such, the project has struggled to capture not only the esoteric textual features of Eaton’s early film manuscripts, but also many of the features now standard to the genre of the screenplay.

This paper will outline some of the standard language and bibliographic features that define the screenplay, its prototypes (such as the photoplay, the continuity, and the scenario), and the various proprietary standards now employed by the screenwriting industry (such as Final Draft XML, Celtx, Adobe Story XML, Fountain, and Fadeln). As we will suggest, Eaton’s various texts—such as continuity scripts, scenarios, and screenplays—serve as a rich example of the complex textual history of the screenplay as a distinct genre that require more nuanced encoding mechanisms than the format-centric proprietary ones listed above can offer, but which the TEI, in its current state, cannot accommodate. Drawing on our work in encoding Eaton’s film texts, we will describe the project’s customizations for encoding screenplays and outline possibilities for refining the TEI’s Performance module to better capture the nature of screenplays as a complex genre of dramatic text in its own right.

## Bibliography

Chapman, Mary, and Sydney Lines (2023). “The first Asian screenwriter in Hollywood’s 1920s ‘dream factory,’ Winnifred Eaton, challenged its racism.” *The Conversation*, 10 April. <https://theconversation.com/the-first-asian-screenwriter-in-hollywoods-1920s-dream-factory-winnifred-eaton-challenged-its-racism-202292>

Johnson, Kevin R. (2019) *The Celluloid Paper Trail: Identification and Description of Twentieth Century Film Scripts*. Oak Knoll Press.

Price, Steven (2013). *A History of the Screenplay*. Palgrave Macmillan.

# Managing Different (Encoding) Cultures when Building a TEI Corpus of Historical German Newspapers

Susanne Haaf, Berlin-Brandenburg Academy of Sciences and Humanities, Germany

Poster

**Keywords:** newspaper corpus, homogeneity, data curation, TEI annotation

## Abstract

### Introduction

Long before the 20th century, reading the newspaper was already a very popular activity and the newspaper was established as a mass medium, leading to its huge impact in various areas, as e.g. language usage and change. Notwithstanding, the Deutsches Textarchiv project (DTA, since 2007) did not have newspapers on the primary list when creating a TEI corpus to document the development of the New High German language between 1600 and 1900. While the project's focus was on printed books from various domains (Geyken, Haaf 2018) the ambitious schedule did not allow for newspapers to be considered for the time being. These hence became important when the DTA started to include digital data from other sources. Then, with the support of many partners a handsome historical TEI newspaper corpus was assembled, which shall be presented as a poster here.

### Aggregated Corpora

The DTA by now comprises 2.039 issues of different historical newspapers dating back to a time period of more than years, from 1609 to 1929, the most voluminous corpora being the Mannheim Corpus of Historical Newspapers (Haaf, Schulz 2014; Fiechter et al. 2019), the Neue Rheinische Zeitung (New Rhenish Newspaper)<sup>[1]</sup>, the Hamburgischer Correspondent (Hamburg Correspondent; Schuster, Wille 2017)<sup>[2]</sup>, the Aviso of 1609<sup>[3]</sup>, and a corpus of issues of the Allgemeine Zeitung from two different sources<sup>[4]</sup>. In addition, several journal corpora have been aggregated by the DTA.

## Coding Cultures

The motto of Coding Cultures suits the historical newspaper corpora of the DTA in different ways:

1. Similar to all data curation for the DTA (Geyken et al. 2018), one challenge was to convert from diverse source formats (following different coding cultures) into one common TEI format, the DTA Base Format (DTABf, since 2011; Haaf, Geyken, Wiegand 2014).
2. Where newspapers were newly digitized (Georgi, Haaf: appearing; Haaf, Schulz 2014; Schuster, Wille 2017), annotations had to suit individual project rules and interests while preserving homogeneity of the corpus.
3. Over the centuries, newspapers had developed increasingly complex layout specifics, following different traditions and resulting in complexity for annotation.

Hence, the challenge was nothing less than to ensure homogeneous TEI annotation in a newspaper corpus that spans several centuries and gathers material from various sources. This was only possible because of a community effort (including willingness to share, help, compromise, and follow standards) in order to create a research resource for everyone. The poster will present the corpus and the challenges of its creation, also raising detailed questions about TEI newspaper markup itself.

## Bibliography

DTA (since 2007): Deutsches Textarchiv. Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache. Prepared by Matthias Boenig, Alexander Geyken, Susanne Haaf, Bryan Jurish, Christian Thomas, and Frank Wiegand. Ed. by Berlin Brandenburg Academy of Sciences and Humanities. Berlin. Online: <https://deutschestextarchiv.de/> (accessed 2023-07-31).

DTABf (since 2011): Deutsches Textarchiv – Basisformat, ed. by Deutsches Textarchiv (DTA) and DTABf Steering Committee (Susanne Haaf, Matthias Boenig, Alexander Geyken, Christian Thomas, Frank Wiegand, Daniel Burkhardt, Stefan Dumont & Martina Gödel). Berlin. Online: <https://deutschestextarchiv.de/doku/basisformat> (accessed 2023-06-22).

Benjamin Fiechter, Susanne Haaf, Amelie Meister, and Oliver Pfefferkorn: *Presseschau um die Jahrhundertwende: Neue historische Zeitungen im DTA*. In Im

Zentrum Sprache. Untersuchungen zur deutschen Sprache, 6 February 2019.  
Online <https://sprache.hypotheses.org/1471> (accessed 2023-08-01).

Alexander Geyken, Matthias Boenig, Susanne Haaf, Bryan Jurish, Christian Thomas, and Frank Wiegand (2018): Das Deutsche Textarchiv als Forschungsplattform für historische Daten in CLARIN. In *Digitale Infrastrukturen für die germanistische Forschung*, ed. by Henning Lobin, Roman Schneider, and Andreas Witt (Germanistische Sprachwissenschaft um 2020 6). Berlin/Boston, pp. 219–248. DOI: 10.1515/9783110538663-011.

Geyken, Alexander and Susanne Haaf (2018): Integration heterogener historischer Textkorpora in das Deutsche Textarchiv. Strategien der Anlagerung und Perspektiven der Nachnutzung. In *Korpuslinguistik*. Ed. by Joachim Gessinger, Angelika Redder and Ulrich Schmitz, with support of Wilfried Stölting (Osnabrücker Beiträge zur Sprachtheorie 92), Duisburg, pp. 175–192.

Georgi, Christopher and Susanne Haaf (appearing): Die Volltextdigitalisierung der „Allgemeinen Zeitung“ (1830–1929). Historischer Hintergrund, Workflow und Forschungsperspektiven. In *Historische Textmuster im Wandel. Neue Wege zu ihrer Erschließung*. Ed. by Susanne Haaf and Britt-Marie Schuster, with support of Frauke Thielert (RGL). Berlin/Boston.

Haaf, Susanne, Alexander Geyken, and Frank Wiegand (2014): The DTA “Base Format”. A TEI Subset for the Compilation of a Large Reference Corpus of Printed Text from Multiple Sources. *Journal of the Text Encoding Initiative (jTEI)* 8. DOI: 10.4000/jtei.1114.

Susanne Haaf, Matthias Schulz: *Historical Newspapers & Journals for the DTA*. In: *Language Resources and Technologies for Processing and Linking Historical Documents and Archives – Deploying Linked Open Data in Cultural Heritage – LRT4HDA*. Proceedings of the workshop, held at the Ninth International Conference on Language Resources and Evaluation (LREC’14), May 26–31, 2014, Reykjavik (Iceland), pp. 50–54.

Schuster, Britt-Marie and Manuel Wille (2017): Die Volltextdigitalisierung der „Staats- und Gelehrten Zeitung des Hamburgischen Unpartheyischen Correspondenten“ und ihrer Vorgänger (1712–1848) und ihr Nutzen. Befunde zur Genese und zum Wandel von Textmustern. In *Die Zeitung als das Medium der neueren Sprachgeschichte? Korpora, Analyse und Wirkung*, ed. by Oliver Pfefferkorn, Jörg Riecke, and Britt-Marie Schuster. Berlin/Boston, pp. 99–119.

## About the author

**Susanne Haaf** holds a degree in German philology and Computational Linguistics (M. A.) from the University of Heidelberg. Currently, she works as a research associate at Berlin-Brandenburg Academy of Sciences and Humanities, where she has been engaged in the projects [DTA](#), [CLARIN-D](#), [t.evo](#) and (till present) [ZDL](#), all of which involved the preparation and maintenance of TEI corpora. She finished and defended her PhD thesis at the University of Paderborn in 2022, which contains work on the computational analysis of patterns which differentiate historical devotional text types.

## Notes

---

1. Online <https://deutschestextarchiv.de/nrhz/> (accessed 2023-08-01). ↵
2. Online <https://www.deutschestextarchiv.de/doku/textquellen#correspondent> (accessed 2023-08-01). ↵
3. Online [https://www.deutschestextarchiv.de/anonym\\_avis\\_1609](https://www.deutschestextarchiv.de/anonym_avis_1609) (accessed 2023-08-01). ↵
4. Online <https://www.deutschestextarchiv.de/doku/textquellen#augsburgerallgemeine>; <https://www.deutschestextarchiv.de/sammlungen/tevo/> (accessed 2023-08-01). ↵

# Materiality in editions of 20th-century paperbound correspondence

**Benjamin W. Bohl**, Goethe-University Frankfurt, Germany

**Dennis Ried**, University of Paderborn

**Patrick Dziurla**, Goethe-University Frankfurt, Germany

**Philipp Kehrer**, Universität für Musik und darstellende Kunst Wien

**Katharina Bleier**, Universität für Musik und darstellende Kunst Wien

**Jannik Franz**, Universität für Musik und darstellende Kunst Wien

**Thomas Kollatz**, Akademie der Wissenschaften und Literatur | Mainz

**Panel**

**Keywords:** Correspondence, Materiality, 20th Century, Critical Edition

## Abstract

### Introduction

Around the mid of the 19th century, new technologies emerged that revolutionised the writing and dispatching of postal documents – the invention of the telegraph or the prevalence of typewriters and copying machines, to name just a few. This shifting media landscape is characterised by a diversity that peaks in the 20th century. Consequently, letter editions are confronted with manifold appearances and manifestations: postcards, letter cards, telegrams, aerogrammes, handwritten letters with pre-printed passages, typescript letters, and many more – all of which might contain handwritten corrections or additions.

The diversity of carrier and writing media, technologies and techniques is a massive challenge for critical editions, not only regarding the contents – the actual message – but also the document itself – the carrier medium. But what is the state of the art when dealing with such material? Does the medium influence the content? How do we adequately handle its materiality? What challenges arise in the process?

The current TEI model for encoding correspondence implements a solid, prototypical content model for a letter body (<opener>, <content>, <closer>). However, the diverging forms of 20th-century correspondence push this model towards its limits, especially when they neglect parts of this prototype (<opener>

or <closer>) or add to it (address blocks, other content such as poems, drawings or written music in extensive postscripts) or are entirely free-form compositions. It is nevertheless desirable to semantically differentiate between formal aspects and the actual content of a letter. While thinking of opener and closer as the disposable parts of a letter might be perfectly true for, e.g., a corpus analysis, informally addressing a friend with his nickname or leading with an austere subject line also transport appreciation, emotion or professional distance.

Moreover, business letterheads, postcards and telegrams introduce new formal (pre-printed) structures, i.e., defined areas for specific information. For example, telegraph forms do not only contain fields for origin, destination and the actual message but also for the documentation of dispatch, transmission and reception of the telegraph message. On the one hand, these constitute an additional layer that might cause structural conflicts with existing TEI models. On the other hand, they constrain the authors when writing their message, thus provoking creativity when available space gets scarce: not only by maximising the density of a message within a limited number of characters (even resulting in a specific 'telegram style' of language) but also by exceeding across the spatial boundaries of a postcard, where text on the illustrated side or on the margins are no rarities.

But what should be considered the primary structure of a corresponding encoding: the logical text sequence or the spatial distribution and the description of the carrier, i.e., the materiality of the document? Is it not the interference of textual layers – the pre-printed text parts, handwritten additions, patched elements (e.g., glued in), stamps etc. – that needs to be addressed and calls for concepts of transcription or even edition of the carrier medium instead of banning its particularities to abstract descriptions?

## **How the panel will be structured**

This panel brings together editors of correspondence from different disciplines. The following projects have stated their active participation:

- Auden Musulin Papers. A Digital Edition of W. H. Auden's Letters to Stella Musulin  
<https://amp.acdh.oeaw.ac.at/>
- The Correspondence between Arnold Schönberg and his Publishers Universal-Edition and Verlag Dreililien. A Critical Hybrid Edition  
<https://www.mdw.ac.at/imi/?PagelId=4284>



- Bernd Alois Zimmermann-Gesamtausgabe. Historisch-kritische Ausgabe seiner Werke, Schriften und Briefe  
<https://www.zimmermann-gesamtausgabe.de>
- Henze-Digital – Hans Werner Henze’s network of artists  
<https://henze-digital.zenmem.de>
- Buber-Korrespondenzen Digital. Das Dialogische Prinzip in Martin Bubers Gelehrten- und Intellektuellennetzwerken im 20. Jahrhundert  
<https://www.adwmainz.de/projekte/buber-korrespondenzen-digital/informationen.html>

After a short general introduction to the discussed field, the intention is to motivate discussion by the panellists and the audience as soon as possible. In order to guide the discussion, the panellists will give short impulse presentations on aspects of the topics described above that will be interleaved by time for general discussion.

We plan to document the event and present the outcomes to facilitate further discussion, e.g., within the SIG Correspondence or the wider TEI community.

## About the authors

**Benjamin W. Bohl** – Goethe University Frankfurt

**Dennis Ried** – University of Paderborn

**Patrick Dziurla** – Goethe University Frankfurt

**Philipp Kehrer** – Universität für Musik und darstellende Kunst Wien

**Katharina Bleier** – Universität für Musik und darstellende Kunst Wien

**Jannik Franz** – Universität für Musik und darstellende Kunst Wien

**Thomas Kollatz** – Akademie der Wissenschaften und Literatur | Mainz

# MEI Basics

Maristella Feustle, University of North Texas, United States of America

Workshop

**Keywords:** pedagogy, basics, beginners

## Abstract

This virtual workshop will proceed similarly to that of the previous two years, and is intended to give absolute beginners extended explanations of and practice with XML basics and rudimentary MEI encoding. The content will include, but not be limited to that of the MEI online tutorials, in order to maximize comprehension and retention of information through repetition and the gradual addition of new information in successive exercises. This workshop will allow for participation by those who are interested in MEI, but cannot travel to Paderborn at this time, and it offers both a smoother transition for users into the content of other conference workshops, as well as a recruiting tool for prospective users who don't know where to begin.

# MEI encoding and analysis of “airs connus” across multiple genres

David A. Day, Brigham Young University, United States of America

Long Paper

**Keywords:** airs connus, comédie-vaudeville, France, music

## Abstract

The term *air connu* may have multiple interpretations in different contexts of 19th-century French music. In its strictest sense it can refer to known melodies that were easily recognizable to audiences of dramatic music including, *opéra-comique*, *ballet-pantomime*, and *comédie-vaudeville*. These melodies may have originated in the popular vernacular or could have achieved recognition as an aria in a successful opera. These *airs connus* were subsequently reintroduced in different genres to recall the original text or meaning to help convey a dramatic plot (as in the case of *ballet-pantomime* where there are no lyrics) or order to establish elements of parody and satire by juxtaposing the original meaning in an altered dramatic context. The term is also used in the context of numerous anthologies that codify the most most popular tunes that were regularly reintroduced in multiple musical-dramatical contexts. Furthermore, the term appears frequently in the context of adaptations of popular melodies arranged for various instruments for amateur consumption.

A better understanding of the extent of *air connu* use across multiple genres will enable a revised assessment of the importance of popular music in 19th-century France. The interconnectivity between the canon of French opera and the more ephemeral productions of the Parisian boulevard theaters is a popular theme in current research. The *air connu* is an essential aspect of this relationship. Furthermore, the use of these same tunes in the vast repertory of amateur arrangements for salon performances is key to understanding the interdependence of all these genres. Arrangement's intended for domestic consumption were no doubt driven by the popularity of successful operas, but also helped reinforce the broad recognition of the same tunes when they appeared in contexts of parody.

The current project aims to build an extensive corpus of MEI encoded *airs connus* and adapt the analytical tools of the CRIM Intervals project (<https://sites.google.com/haverford.edu/crim-project/home>) to identify uses of these melodies across multiple genres including opera, ballet, French vaudeville, amateur arrangements, and *recueil* of popular airs. The corpus is being developed through MuseScore input and conversion to MEI. OMR using SmartScore is also being tested. The initial production focus is the sixth edition of the *La clé du caveau* and aria melodies found in the *répétiteur* or *chant* parts from the *ballet-pantomime* and *comédie-vaudeville* repertoire of the Théâtre de la Monnaie housed at the Archives de la Ville de Bruxelles. A project to photograph all of the orchestral material for *comédie-vaudeville* (including *répétiteur* and *chant* parts) from this important collection was completed July 2023. Melodic incipits for the most important *opéra-comique* found in RISM are also being exported as MEI files. To these corpora will be added an existing substantial body of more than 5,000 MEI encoded melodies from popular tunes arranged for the harp and drawn from BYU's extensive International Harp Archives.

This presentation will review the methodology and goals of the project. A report on the current number of files created and use of the CRIM Intervals analysis tools will also be provided. Future work on the texts of *airs connus* using TEI will be assessed.

## About the author

**David A. Day** has worked as Curator of Music Special Collections at Brigham Young University since 1986. He is an adjunct professor and teaches the graduate research methodology course at the BYU School of Music. He was awarded the PhD in musicology from New York University in 2008. He is currently building a MEI corpus of French *airs connus* or known melodies.

# MEI Transformation for the Winterreise Dataset

Joshua Neumann, Akademie der Wissenschaften und der Literatur Mainz, Germany

Poster

**Keywords:** Data Standardization, Audio, Performance Analysis, Texted Music

## Abstract

Just as Franz Schubert is responsible for the mature development of the song cycle, so too is Dietrich Fischer-Dieskau for enhancing the genre's popularity from the mid-twentieth century onward. In addition to numerous recorded performances, Fischer-Dieskau also collaborated on a critical edition of *Die Winterreise*, situating the cycle as unified, complete 'work.' As Natasha Loges has shown, however, performing this cycle in its complete form was not nineteenth-century convention. Moreover, recordings as late as 1943 eliminate strophes of some songs to accommodate limitations of recording technologies. While these reasons for the cycle's incompleteness differ, the fact remains that presenting it in truncated form was common practice until the mid-1940s.

Dietrich Fischer-Dieskau's longstanding relationship with Franz Schubert's *Die Winterreise* resulted in no fewer than fifteen recordings in the period 1948–1990. So many recordings over such a long time is exceptionally rare for any singer, given the effects of age upon a singer's choice of repertoire. Fischer-Dieskau's reputation as a singer emphasizes his ability to communicate text and music effectively, as both individual and fused elements, as Alan Blyth summarized his singing for *Oxford Music Online*. "His interpretations set standards by which other performances were judged. They were based on command of rhythm, a perfect marriage of tone and words, an almost flawless technique and an unerring ability to impart the right color and nuance to a phrase." Given Fischer-Dieskau's importance in performance practice history, and the reverence in which singers, pedagogues, and performance studies scholars hold him, understanding how his performances of *Die Winterreise* evolved is imperative for understanding the cycle's ongoing creative history.

Such an expansive body of recordings, paired together with such a distinctive status, invites a musicologically-oriented corpus study focused on audio data in

relationship to both music and text, and to/with which interdisciplinary audiences can interact. Two extant projects offer a foundation upon which such an interdisciplinary analysis is possible: the *Schubert Winterreise Dataset*, and the *Music Performance Markup* toolkit. The Winterreise dataset is a notable and valuable contribution to open science for engaging with this musical content, even as it requires further transformation for use within the MEI environment. Processing the XML requires data quality review and schema transformation into MEI. This workflow will produce diplomatic editions based on the sources used in compiling the Winterreise dataset. Having this song cycle in MEI will enable its use with the Musical Performance Markup infrastructure.

This poster overviews this transformation, situating it in relationship to storing audio markup data along three axes: tempo, dynamics, and timbre in relationship to musical and textual content. Doing so also sets out a technological workflow for responding to the broader musicological questions surrounding the relationship Dietrich Fischer-Dieskau had with *Die Winterreise* over a half century. Such a workflow can in turn be useful for future MEI-based considerations of musical performance and its resulting audio data.

## Bibliography

Berndt, Axel. "Music Performance Markup: Format and Software Tools Report." In *Music Encoding Conference Proceedings 2021* (pp. 57–63). Humanities Commons, 2022. DOI: <https://doi.org/10.17613/s357-e217>.

Blyth, Alan. "Fischer-Dieskau, Dietrich." *Grove Music Online*. 2001; Accessed 20 Jul. 2023.  
<https://www.oxfordmusiconline.com/grovemusic/view/10.1093/gmo/9781561592630.001.0001/omo-9781561592630-e-0000009737>.

Loges, Natasha. "Detours on a Winter's Journey: Schubert's *Winterreise* in Nineteenth-Century Concerts." *Journal of the American Musicological Society* 74, no. 1 (2021): 1–42. <https://doi.org/10.1525/jams.2021.74.1.1>.

Weiß, Christof, Frank Zalkow, Vlora Arifi-Müller, Meinard Müller, Hendrik Vincent Koops, Anja Volk, and Harald G. Grohgan. "Schubert Winterreise Dataset." *Journal on Computing and Cultural Heritage* 14, no. 2 (2021): 1–18.  
<https://doi.org/10.1145/3429743>.

“Willkommen Auf Schubert-Online – Der Größten Digitalen Sammlung von Schubert-Autographen, Erst- Und Frühdrucken!” SCHUBERT online. Accessed July 15, 2023. <https://schubert-online.at/activpage/index.php>.

# mei-friend v1.0: Music Encoding in the Browser

Werner Goebel, mdw – University of Music and Performing Arts Vienna, Austria

David M. Weigl, mdw – University of Music and Performing Arts Vienna, Austria

Long Paper

**Keywords:** mei-friend, Web application, Music encoding, MEI, editor

## Abstract

Music encoding is an inherently interdisciplinary endeavour, bridging technological and musical scholarly perspectives, usually through collaboration of individuals from distinct and divergent academic backgrounds. We present mei-friend, a Web application that smooths the ground for such collaboration by simplifying and speeding up music encoding workflows, while flattening the difficulty curve experienced by those newly entering the field.

Originally developed as a plug-in for the Atom text editor, mei-friend has been redesigned and extended as an open-source, Web-native browser application, written in vanilla JavaScript. Verovio provides its music engraving engine, and CodeMirror its XML text editor. The application integrates important community resources, including the MEI schemas for autocompletion and validation, the MEI Guidelines for documentation, GitHub and Git Lab for collaborative encoding, and Solid for decentralised annotation sharing. Various music encoding formats may be imported into the application, from the local file system (optionally by drag-n-drop), from a Web location specified via URL, or through the Git cloud service integration. Edited encodings may be exported as rendered SVG, PDF, or MIDI. External sources – e.g., community tutorials, Git issues, Web discussion boards, or scholarly articles – may use hyperlinks with specified parameters to link into mei-friend, directing it to fork a given Git repository, open a specified encoding, flip to a particular page, or to select a set of musical elements, enhancing the application's support for shared discussion and collaborative editing. Stand-off annotations employing the W3C Web Annotation Data Model and associated vocabularies may be authored and shared, supporting data-level integration with other Web applications.



Though guided by practical encoding requirements of the digital music research projects that have funded its development, the application has been designed with a firm eye on pedagogy, lowering barriers to entry and providing ease-of-use. Considerations include the removal of any installation or specific system requirements – all that is needed is a Web browser. Each editing function is available through graphical menus available in a growing number of languages, as well as via keyboard shortcuts, reducing language barriers and aiding discoverability while supporting optimised workflows for more experienced users. A dedicated ‘mei-friend Help’ website provides documentation of all basic and advanced functionalities, and describes how to contribute to mei-friend’s development. Beside the immediate visual feedback provided by the tightly-coupled integration of XML text and rendered notation, auditory feedback is provided by a built-in MIDI player, further facilitating the verification of encodings by novice users.

In this paper, we describe the application’s implementation and its functionalities, detailing design considerations and surfacing optimizations and extensions compared to the original Atom plugin described at the Music Encoding Conference 2021. We provide a practical evaluation reporting on experiences using mei-friend within a number of music-edition and research projects, and as a pedagogical tool used in MEI tutorials at conferences, project meetings, and in class-room use at different institutions. We conclude with an overview of current limitations, plans for future development, and a call for contributions to mei-friend’s development by community members.

## **Acknowledgements**

Development of the mei-friend Web application is undertaken as part of the Signature Sound Vienna Project, funded by the Austrian Science Fund (FWF) P 34664-G; and supported by E-LAUTE, an international Weave Research (FWF/DFG/SNSF)-funded project (I 6019). The mei-friend Atom plugin package was developed as part of TROMPA, with funding from the European Union’s Horizon 2020 research and innovation programme H2020-EU.3.6.3.1. under grant agreement No 770376. We thank members of the MEI community, particularly Anna Plaksin, Thomas Weber, Oleksii Sapov, David Lewis, and Laurent Pugin for their contributions, feedback and suggestions.

# Meta-Models Matter – The tscore R&D Project and “Die Kunst der Fuge”

Markus Lepper, semantics gGmbH, Germany

Baltasar Trancón y Widemann, semantics gGmbH, Germany; Nordakademie  
Elmshorn, Germany

Poster

**Keywords:** Meta-meta-model, Modeling theory,  
Java, Die Kunst der Fuge, Domain  
Specific Language

## Abstract

The fundamental analysis by Lepper (2021) shows that for precisely defining any music notation or encoding system (a) a clean separation of the syntactic sphere from the semantic sphere is necessary, based on which (b) mathematical quadrupels must be defined which cover the structures of both spheres plus the mappings between them, in both directions. But each practical notation/encoding system has turned out (c) as a pipeline or even as a branching network between multiple such spheres.

When choosing the data for digital encoding, often we cut somewhere into the middle of some larger evaluation pipeline. For instance, explicitly storing spelled pitches ("SPCs" according to Hentschel (2021), like "e flat" and "e natural") skips non-trivial steps of traditional reading, namely computing the combined effect of key signature, sequences of "white key" note heads ("GPCs"), local accidentals, and rules for scoping and priority.

In computer science, encoded objects are often called *models*. Their type definition is called a *meta-model*. The interpretation of any model is only sensible based on its meta-model, which gives the rules for its syntactic and semantic interpretation. Therefore modeling music requires the construction of an appropriate meta-model.

At first glance, this seems a tedious obligation, which in practice often is replaced by "common sense" or "obvious obviousness". But it also is a chance: Making the meta-model explicit for a family of models allows

(a) to check syntax and semantics right from the beginning of the construction,

- (b) to reduce all input, edit, and control efforts to the really necessary,
- (c) to disambiguate the interpretation of the applied notation,
- (d) to re-use the encoded data with a different focus of modeling, by switching the meta-model,
- (e) to extend any notation as needed into a new meta-model without conflicts with other authors and styles, and, last not least,
- (f) to combine and compare models from unrelated sources without the need of an a priori agreement about the encoding format.

All this requires "only" that meta-models can be compared and combined, and that models can be converted between meta-models by applying their forward and backward mappings. To automate this as far as possible, computer engineering for one particular realm always aims at one particular *meta-meta-model* to define the meta-models.

Contrarily, Raimond-Adbdallah (2007) and Fields et.al (2011) define the formats for particular music performance and structure data: thus they define a meta-model for these models. The meta-meta-model employed by both is a general purpose one, namely Web Ontology Language (OWL).

Lepper (2021) has shown that in the realm of mere conventional usage of Common Western Notation (CWN) more than eight hundred meta-model parameters are relevant. The strategy to find one single meta-model covering all these variants (up to 6.6E240 !) appears unfeasible, esp. if not only the syntax but also the semantics shall be meta-modelled precisely.

The authors' "tscore" project thus takes the different approach of constructing a dedicated meta-meta-model into which particular meta-models can be plugged in as needed (Lepper 2013). It realizes a generic input format similar to the conventional orchestra score, with freely definable time model (currently restricted to one global) and arbitrarily mixed parameter tracks, independent for each voice.

The approach is co-algebraic: The self-identity of each "notation event" (according to Lepper (2021)) is axiomatic, all parameters are just accidental. This allows to encode any phase of the above-mentioned pipelines, from concrete front-end syntax to pure semantic data (late states in traditional notation evaluation).

The meta-model ingredients are currently combined using the Java programming language plus Embedded Domain Specific Languages (EDSLs = library API calls or library classes, which to combine means to do programming).

The implementation is open source and in the public domain. It currently provides meta-models for experimental graphic composition (tscore 2020), and a family of meta-models for conventional Common Western Notation, with several translation algorithms into (subsets of) musicXML, lilyPond, MEI, etc. (Lepper-Trancón 2019). The EDSLs are carefully designed to support the usage by domain experts (non-computer-language experts), employing sophisticated algorithms for error diagnosis (Lepper-Trancón 2018).

Thanks to property (b) from above, the complete "Die Kunst der Fuge" has been entered in only three days. The co-algebraic Java model is now in the public domain and can be used for further transformation, analysis, and interpretation.

## Bibliography

Fields, Ben, Kevin Page, David De Roure, and Tim Crawford (2011) *The Segment Ontology: Bridging Music-Generic and Domain-Specific*. ICME 2011.

<https://eprints.soton.ac.uk/272698/1/admire2011.pdf>

Hentschel, Johannes, Fabian C. Moss, Andrew McLeod, Markus Neuwirth, and Martin Rohrmeier (2021) *Towards a Unified Model of Chords in Western Harmony*, in: Proceedings of the Music Encoding Conference MEC 2021, <https://music-encoding.org/conference/proceedings.html>

Lepper, Markus (2021) *de Linguis Musicam Notare*. epOs, Osnabrück.

Lepper, Markus and Baltasar Trancón y Widemann (2013) *tscore: Makes Computers and Humans Talk About Time*, in: Proceedings of KEOD 2013, 5th Int. Conference on Knowledge Engineering and Ontology Development, Instincc, Portugal, 2013, p. 176-183, doi:10.5220/0004540701760183.

Lepper, Markus and Baltasar Trancón y Widemann (2019) *metricSplit: An Automated Notation of Rhythm Aligned with Metric Structure*, in: Journal of Mathematics and Music, Taylor Francis, London.

Lepper, Markus and Baltasar Trancón y Widemann (2018) *Translets --- Parsing Diagnosis in Small DSLs, with Permutation Combinator and Epsilon Productions*, in:

Tagungsband des 35ten Jahrestreffens der GI-Fachgruppe "Programmiersprachen und Rechenkonzepte", IFI Reports, Vol. 482, University of Oslo, Oslo, 2018, p. 114-129.

Raimond, Yves and Samer Abdallah (2007) *The Timeline Ontology*

<https://motools.sourceforge.net/timeline/timeline.html>

tscore (2020) *Example Instances of the TScore Project Infrastructure*

<http://markuslepper.eu/sempart/tscoreInstances.html>

## About the authors

**Markus Lepper**, composer, music theorist, and computer scientist. He holds a PhD each in computer science and musicology. Works and lives in Berlin and is co-founder of semantics gGmbH Berlin.

**Baltasar Trancón Widemann** holds a PhD in computer science from TU Berlin and a Habilitation degree from the University of Bayreuth. He has worked as a researcher in academia and as an industrial software engineer. He is currently professor of programming at Nordakademie Elmshorn, and co-founder of semantics gGmbH Berlin.

# Metadata in the DFG Viewer for Musical Sources: Requirements for the exchange of meta and structural data in MEI format

Irmlind Capelle, Universität Paderborn, Germany

Kristina Richts-Matthaei, Universität Paderborn, Germany

Poster

**Keywords:** Metadata, data exchange, METS/MODS

## Abstract

The DFG Viewer is a browser web service for displaying digitized material from decentralized library repositories. It previously had an OAI-PMH interface for exchanging meta and structural data in the METS/MODS and METS/TEI format. With the help of this data, the display of a digital record is enriched and, if necessary, supplemented with additional functions. Based on the MODS or TEI data, extended metadata up to detailed source descriptions can be displayed, while with the help of the METS data, access to the digital record is structured according to formal and/or content criteria.

In a two-year project, this offer was extended for musical sources, also enabling the integration of MEI or MusicXML files with the help of Verovio as well as the integration of audio files. In addition, the presentation of sources with bar-by-bar access and the parallel display of voices, among other things, were implemented.

To enable the exchange between research data and source presentation in the long term, some details have to be considered when mapping the metadata between METS/MODS and MEI, especially since the MODS format does not take FRBR into account.

The poster will present both the recommendations for metadata acquisition in MODS that have been developed for musical sources and their equivalent in MEI.

# Mixing Modalities: Graphical and Text-Based Interaction in Music Notation Editing

**Matthias Nowakowski**, Hochschule für Musik Detmold, Germany

**Axel Berndt**, Technische Hochschule Ostwestfalen-Lippe

**Anna Plaksin**, Johannes Gutenberg-University Mainz, Germany

**Nevin Şahin**, Hacettepe University Ankara State Conservatory

**Aristotelis Hadjakos**, Hochschule für Musik Detmold, Germany

**Poster**

**Keywords:** music notation, graphic modality, user interface design

## Abstract

Editing of music notation is done in graphical as well as in text-based interfaces. The latter is still common practice in the MEI community while the prior tends to be the more intuitive approach. In the context of the currently developed research and editing Tools mei-friend (Goebel, Weigl, 2022) and VIBE (Nowakowski, Hadjakos, Stärk, 2022) we discuss possibilities to combine both worlds by enriching text-based music editors with specialized GUI widgets. The aim is not a WYSIWYG interface of the whole music notation but lightweight, context-sensitive graphical elements that complement the XML encoding.

Music notation is complex. Due to the high information density of music notation, it is clear for a skilled reader to determine an element in its context and how it is to be played. But often there is room for interpretation, which arises e.g. by historical distance or changing purposes and contexts (Plaksin et al., 2022). So how does one map this complex sign system on machine interactions?

Interaction with music notation usually happens through a mixed modality of mouse and keyboard, the latter of which is highly optimized towards editing textual information in a one-dimensional way. Common music notation programs use these modalities to accomplish their manifold tasks. Formats like MusicXML or MEI can textually represent this information density, complexity and hierarchy in which the element is placed. A graphical method would have the advantage to translate this

information density more quickly into text by a unique gesture or interaction with the user interface.



Some examples could be:

- Preview of single MEI Elements while typing.
- Interactive navigation from text to musical score and vice versa.
- Templating and recommendations while writing text and music
- Versioning overviews focussed on distinct symbols or XML sequences.

In an effort to get in touch with a diverse audience to understand the individual patterns of interaction with graphic modalities, we ask these open questions:

- What is the main interaction medium with music notation?
- Which kind of information and visual aids do different target groups need?
- How can an interface support the understanding of complex sign systems and its relation to the XML encoding?
- How can the interaction with multiple textual layers be supported during the editing of music?
- What visual aids could editors need to edit XML more quickly?
- What is the main interaction medium (XML, music graphics, LilyPond, ...)?
- Which kind of information do different target groups need?
- As users need to understand how XML and music notation correspond: How can an interface support learnability in this respect?

Participants of the conference will have the opportunity to try out both editors and leave ideas for widgets on the poster which then will be discussed in the following paper.

## Bibliography

Goebel, W.; and M. Weigl, D. Alleviating the Last Mile of Encoding: The *mei-friend* Package for the Atom Text Editor. In Münnich, S.; and Rizo, D., editor(s), *Music Encoding Conference Proceedings 2021*, pages 31–39, 2022. Humanities Commons <http://doi.org/10.17613/45ag-v044>.

Nowakowski, M., Hadjakos, A and Stärk, A, Toward Interactive Music Notation for Learning Management Systems, in: *International Journal on Innovations in Online Education*, vol. 5, iss. 3, 2022. Begel House Inc.

Plaksin, Anna, David Lewis, Nevin Şahin, and Axel Berndt, 'Sharing MEI: Common Semantics in Diverse Musics?', in: *Music Encoding Conference* (Halifax, Canada: Dalhousie University, Music Encoding Initiative, 2022).

## About the authors

**Matthias Nowakowski** studied Musicology and Philosophy at the University of Cologne (Germany), and Mediainformatics at the University of Applied Science in Düsseldorf (Germany). After his thesis about automatic transcription for electroacoustic music he concentrated his academic work on music information retrieval and is currently working at the Center of Music and Film Informatics in Detmold (Germany). He is currently writing his PhD thesis about interaction and user experience with digital musical notations.

**Axel Berndt** is a computer scientist with a musical background. He works currently as a postdoctoral researcher at the Center of Music and Film Informatics in Detmold, Germany. His research includes various topics in the music technology field, such as musical human-computer interaction, music performance modeling and analysis, game scoring, and generative techniques.

**Anna Plaksin** is a postdoctoral research associate at the Institute of Art History and Musicology (IKM) at Johannes Gutenberg University Mainz. She currently works on support for editorial markup in the *mei-friend* Web application.

**Nevin Şahin** is an assistant professor of music theories at Hacettepe University Ankara State Conservatory. Focusing on non-western music traditions which utilize different notation systems such as Hampartsum notation and Byzantine neumes, she tries to bring together MEI and early music research.

**Aristotelis Hadjakos** is a professor of music informatics and co-director of the Center of Music and Film Informatics. He conducts research in the area of musical human-machine interaction. His specific research interests are digital scores, digital humanities, and tangible musical interfaces.

# Multicultural & Multilingual TEI Encoding: A Comparative Study of British Romanticism and Chinese Tang Poetry through the Lens of Romantic Ecology and Affect Narratology

Alisa C Chen, Northeastern University, MA, USA

Short Paper

**Keywords:** Sentiment analysis, Data visualization, Romantic ecology, affect narratology, Tang Poetry

## Abstract

By analyzing British Romantic and Chinese Tang poetry, I argue the two periods should be central rather than tangential to existing scholarship and literature on eco-criticism, pastoral poetics, and affect narratology. In addition to analyzing and visualizing the poetry of these two periods through TEI text encoding, my project also aims to demonstrate the beauty of the multilingual and multicultural linguistic aspects of British and Chinese cultural influences. Both Romantic and Tang poetry explore human consciousness and translate emotional interiority into written language. Therefore, both poetic traditions rely heavily on eco-imagery and prosopopeia to garner deeper reader response and engagement. Both poetic traditions are indicative of each countries' respective national ideological history, which is made evident through the effective usage of eco-imagery, colloquial language, and performativity.

For this project, I've created a WordPress website that houses a curated collection of twenty poems (ten from each period). The Tang poems are in traditional Mandarin and translated by Xu Yuanchong. Features of the website include: Side-by-side modules of the Romantic and the English-translations of the Tang poems; Hovering features that note pastoral elements, poetic devices, emotions, and spatiality; Background of the authors, their famous works, and signature features of their writing; Two network visualization of the poets.

I use TEI schema (XML) and Roma customization to encode the poetry, OpenRefine to clean my data, Cytoscape to visualize the network of the poets, CSS to visualize my encodings on the WordPress website, and Wave to check for accessibility.

The website provides an intersectional analysis of the Romantic and Tang poems that embraces the nuances of Eastern and Western poetic traditions. I hope my project will bring scholars and communities that study poetry of both the British Romantic and Chinese Tang Dynasty traditions together and further inspire interdisciplinary collaboration through digital literary means.

## **About the author**

**Alisa C. Chen** is an English M.A. student at Northeastern University. She is pursuing Graduate Certificates in DH and WGSS. She studies Chinese Tang and British Romanticism poems with interests in Romantic ecology, affect theory, narratology, the anthropocene, and aesthetics. She is working on the Thoreau Journal Drawings website and her DH project.

# Music-Text Interlinking as a Challenge for Digital Encodings of Music-Theoretical Writings

Torsten Roeder, Universität Würzburg, Germany

Fabian C. Moss, Universität Würzburg, Germany

Maik Köster, Universität zu Köln, Germany

Long Paper

**Keywords:** music encoding, text encoding, digital humanities, music theory

## Abstract

Music-theoretical writings are often multimodal, in the sense that they may contain textual as well as musical elements, and even diagrams or images. These different modalities are moreover frequently related to one another. This can happen, for instance, when specific music examples are provided to illustrate music-theoretical concepts that are discussed in the text, or when certain passages from “real” compositions are analyzed or commented on. In addition, it can occur that musical characters and symbols such as notes, rests, or clefs, are not separated from but rather embedded within the textual flow. In general, multimodalities of text and notated music can also be found in numerous other types of sources, such as textbooks, newspapers, journals, and letters, but with a remarkable density and variety in music-critical or music-theoretical treatises.

It has been demonstrated that text-music-linking can operate very effectively on a presentational level for text that repeatedly and explicitly refers to sections of a piece of music in standard notation.<sup>[1]</sup> Yet, another challenge for digital encoding are symbols for harmonic or other analyses that are idiosyncratic to only one or a few sources but not common enough to be considered by current music encoding standards and guidelines. This tightly-knit texture of musical and textual interdependence renders music-theoretical treatises a prime case study for digital encodings of documents that partake in both domains. A successful multimodal encoding with effective bidirectional interlinking (music to text and text to music) would moreover facilitate subsequent computational analyses, such as querying for specific terms and concepts or named-entity recognition across documents.

Our contribution presents such a case, namely a corpus of selected 19th-century German music monographs encoded in TEI and Kern (convertible into MEI).<sup>[2]</sup> The selection was based on the relevance of the texts for the so-called harmonic dualism, a topic of at times heated debate.<sup>[3][4][5]</sup> Rather than discussing the content of these works, we here focus on the challenges posed through interlinkage of music and text in TEI/MEI encoded documents. While the basic scenario will consist in references between TEI and MEI portions, text is often part of musical notation itself, which would require text-music linking only within MEI. On the basis of various examples from the corpus, we demonstrate current encoding challenges, and propose and discuss several directions towards their resolution.

## About the authors

**Torsten Roeder**, Universität Würzburg

**Fabian C. Moss**, Universität Würzburg

**Maik Köster**, Universität zu Köln

## Notes

---

1. Klaus Rettinghaus: *Digitale und kritische Edition der 24 Probstücke der Ober-Classe aus Johann Matthesons Großer Generalbass-Schule*, <http://probstuecke-digital.de/>. ↵
2. Text corpus: <https://github.com/DCMLab/ddd/tree/main/data> and music examples: <https://osf.io/qm9z5/>. ↵
3. Cfr. 'Digitizing the Dualism Debate' (2021), <https://dcmlab.github.io/ddd/>. ↵
4. Jorgenson, D. (1963). 'A Résumé of Harmonic Dualism'. *Music & Letters*, 44(1), 31–42. ↵
5. Snyder, J. L. (1980). 'Harmonic Dualism and the Origin of the Minor Triad'. *Indiana Theory Review*, 4(1), 45–78. ↵

# Networking Editions: TEI and the Semantic Web with TAG and RDF

**Elena Chestnova**, Università della Svizzera italiana, Switzerland

**Andreas Kuczera**, University of Applied Science, Gießen, Germany

**Stefan Armbruster**, independent

**Balduin Landolt**, Swiss National Data and Service Center for Humanities DaSCH

**Panel**

**Keywords:** Semantic Web, Text-as-Graph, RDF

## Abstract

Making text and metadata of digital editions addressable and linkable is a task that several recent and current projects in the field of scholarly editing have focused on (e.g. Skaldic poetry projects, the LEAF writer and other LINCS projects, Unlocking Digital Texts, and others). These projects have been exploring ways to connect editions data with other resources on the web and associated issues: re-usability and sustainability of editions, interoperability of text, collaboration across fields and disciplines, and development of federated tools. Discussions around these topics have led to the development of a range of approaches and this panel will give insight into two of these: implementation of TEI in Text-as-graph and incorporation of TEI in an RDF triple store. The panelists will present experiences from real applications in the context of edition projects and a national infrastructure for humanities data.

The panel has been assembled as part of an ongoing project “Semantic TEI” which explores possible pathways towards a semantic ontology to bridge between TEI and existing widely-used ontologies. This is to be developed with a view to emergent potential for future applications and research methods and based on the needs of the research community. Understanding these needs is the main motivation of this panel. Its purpose is to create an opportunity for members of the TEI community to learn about recent work on TEI in the context of TAG and RDF and to question and discuss this work.

## 1. Andreas Kuczera: TEI in Graph-based digital scholarly editions

TEI is a quasi standard to express text and its meaning in digital scholarly editions. The main questions in my statement are:

- Does TEI need the structure of XML to keep the semantic and if yes are there examples for that ?
- Are there ways to structure textual information beyond the approach of Microsoft Word (which works with Layout elements) or TEI ?
- What could this new structure look like ?

One possible way to make further steps in these directions could be the connection between TEI semantics and graph based modeling approaches for textual information.

## **2. Stefan Armbruster: Conversion of TEI-XML to LPG and TAG**

Given my long term background in graph data modeling I am specifically interested in the graph data model used for imported TEI data. The data model is driven by its usage patterns (aka queries) and its extensibility for future use cases.

Some of these additional use cases I can think of:

- adding/querying provenance information
- more data driven / machine accessible model
- Versioning of LPG
- Another interesting aspect is long-time archival of LPG and ensuring it's still accessible decades in the future
- Applying classic data science or graph data science on LPG model

## **3. Balduin Landolt: TEI as RDF: Opportunities, Limitations and Useability**

TEI is normally thought of as an XML specification, however it can also be understood as a general vocabulary serving as a formal ontology independent of any serialization format.



Stand-off-based approaches to text encoding can address some of the issues following the hierarchical structure of XML, and especially lend themselves to graph representations of text. Such graphs can be defined by a TEI ontology.

At the Swiss National Data and Service Center for the Humanities (DaSCH) we have gathered experience in using RDF graphs to represent text with stand-off markup, and are planning to expand this technology to be fully TEI compatible in the future. This approach has shown promise in terms of searchability of markup and its linkage applying semantic web technologies. However, it has also shown problems, specifically with regard to its useability.

Based on this experience, I plan to outline, how a "TEI as RDF" solution might be designed, so that it reaps as many of the benefits of stand-off mark-up and Linked Open Data technologies as possible, while trying to mitigate its inherent drawbacks.

## About the authors

**Elena Chestnova** (Università della Svizzera italiana), chair of the session, is a researcher with the Institute for History and Theory of Art and Architecture, DH lead for the Semper Edition and project lead for the *Semantic TEI* project.

**Andreas Kuczera** is professor for applied digital methodology in the humanities and social sciences at the university of applied science Gießen. He works on graph technologies in the digital humanities and graph based digital scholarly editions.

**Stefan Armbruster** is a freelancer providing consulting services in the field of graph databases based on his 10+ years of experience.

**Balduin Landolt** is a software engineer at DaSCH. He is currently planning a PhD project in Scandinavian Studies/Digital Humanities at the University of Basel.

# Notation editor for transcribing Hindustani classical music Bandish

Kiran Ramakant Karkera, Independent, Germany

Short Paper

**Keywords:** Hindustani Bandish Notation, Sargam notation software

## Abstract

In this paper we present the Bhatkhande Bandish editor, a notation editor for transcribing Hindustani (North Indian classical music) compositions known as *Bandish* in a traditional notation system known as *Bhatkhande* notation.

## Introduction

Hindustani classical music (HCM) is a traditional form of music practiced in the Indian subcontinent. One of the presentation forms in HCM is called *Bandish* (Ranade n.d.), which can be described as a composition based on a set of notes (called a *Raga*<sup>[1]</sup>), in a musical meter (called *Taal*<sup>[2]</sup>). In this paper, we describe a web-based notation editor capable of transcribing and rendering a *Bandish* in a traditional notation system, named the *Bhatkhande* notation system (Schooling n.d.) after its author V N Bhatkhande<sup>[3]</sup>.

This editor is an open-source alternative to proprietary notation editors (Sawant n.d.) (SwarClassical n.d.). It offers two artifacts:

- An open file format specification to store a Bandish document.
- An open-source engine to render a Bandish on a web page.

Additionally, making the project open source allows community contributions for *Raga* and *Taal* metadata which makes it easier to support new *Ragas* and *Taals*. It also makes it possible to support new Indic languages to display the notations in.

## Notation Schema

Top level elements

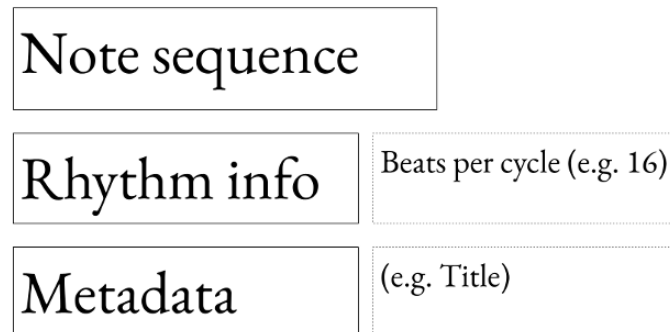


Fig. 1: Top level elements in Bandish document

The notation schema consists of three top level elements:

- A sequence of notes
- Rhythm information
- Metadata such as title.

Each element in the sequence of notes represents musical artifacts that must be played in one beat. This includes:

- One or more notes (called *Svaras*<sup>[4]</sup> in HCM)
- One or more note ornamentations (such as trills or glissando)
- The lyrics to be sung in the duration of one beat.

One beat in the rhythmic (Taal) cycle

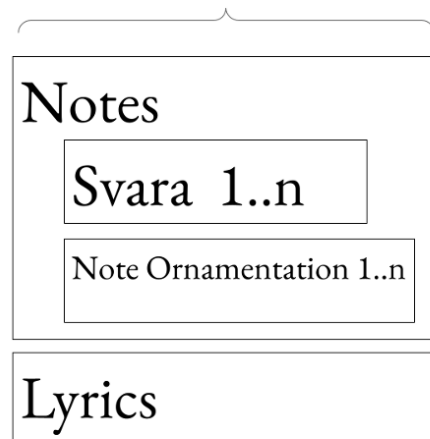


Fig. 2: Information per beat

Each *Svara* is represented so:

Since HCM uses relative note positions called *Svaras* (as opposed to notes defined by a particular frequency), we denote the 12 note positions by the first character of its name (e.g., *s r g m*, like *solfege*).

Raga music is largely (Indian Oral Tradition n.d.) and most compositions are restricted to 3 octaves. Therefore, it is sufficient to restrict the number of octaves to three.

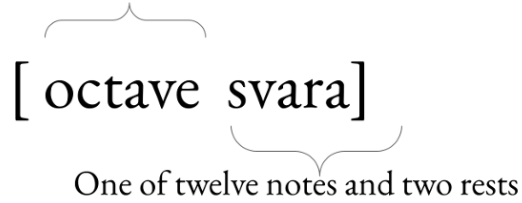
We can thus indicate each note as a vector of two elements.

- The first element indicates the low / medium / high octaves.
- The second indicates the *Svara* or note position.

Each of the *Svaras* is defined thus:

### Svara definition

One of three: low / medium/high



s r g m p d n	7 Svaras
-r -g m+ -d -n	5 flats & sharps
- a	2 Rests

Fig. 3: Svara definition

The Bandish document format is defined using Clojure Spec<sup>[5]</sup> and can be serialized as Transit<sup>[6]</sup> or JSON files. The detailed schema description can be found at the GitHub page for sargam-spec<sup>[7]</sup>.

### Rendering Engine

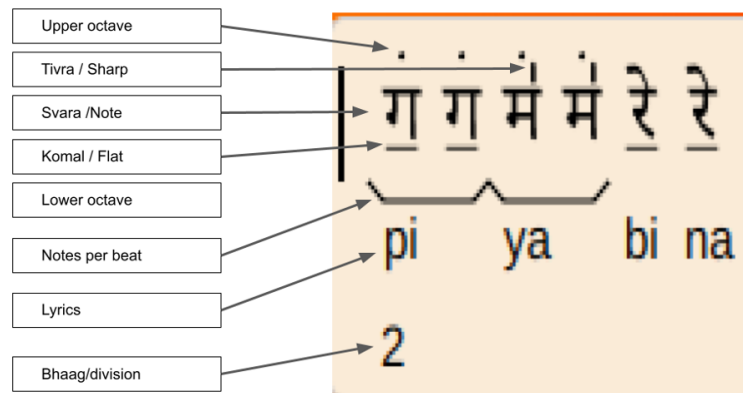


Fig. 4: Vertical orientation of symbols in a sequence of **Svaras** (notes)

Displaying a *Svara* (note) requires the presentation of the following visual elements for the notes:

- The base note or *Svara* as a character in an Indic language.
- Indicating if a note is sharp (*Tivra*) or flat (*Komal*).
- Octave information

The following visual elements indicate the rhythm and timing.

- The number of notes in a single beat
- The position of the note in a rhythmic cycle

Lastly, the lyrics to be sung on the beat are also displayed.

Ideally, a *Svara* (note) should be rendered with a Unicode character. Due to difficulties in combining multiple diacritics in Unicode, the authors chose to use images to render Svaras, requiring 36 images per Indic script.

## Transcribing notations

The mobile viewport was selected as the primary interface to the notation editor. The interface design mimics the look and feel of a messaging application because users may find the interface familiar in order to transcribe notations.

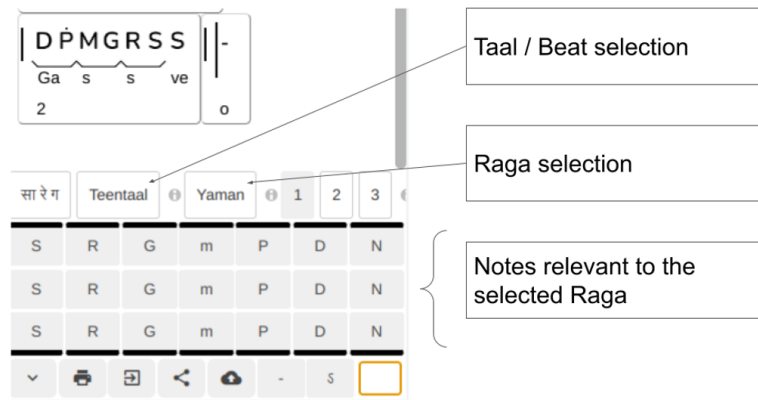


Fig. 5: Interface design that mimics mobile messaging application UI.

## Comparing the rendering to traditional typesetting

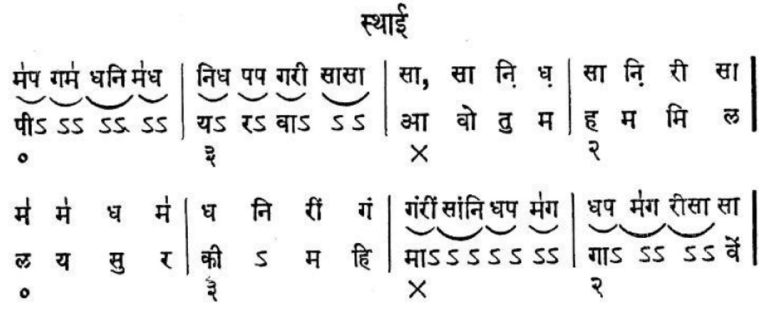


Fig. 6: Excerpt of the Bandish Piyarava from the book *AnoopRagaVilas*



Fig. 7: Reproduction of 'Piyarava' in Bandish editor's rendering engine

Fig 6 shows a *Bandish* that is typeset in the book *AnoopRagaVilas* Vol one (Gandharva 1965). Fig 7 shows the same *Bandish* rendered using the notation editor (Karkera n.d.). Observe that the *Bandish* editor closely mimics the typeset conventions used for publishing *Bhatkhande* notations.

## Conclusions

In this paper we demonstrated a Bandish editor capable of editing and rendering notations for HCM known as Bhatkhande notation. We also presented a file format to store Bandish notation documents.

## Acknowledgements

I would like to acknowledge the Clojure<sup>[8]</sup> community for their generous contribution to the Clojure language in which the Bandish editor application is

implemented. I am also grateful to the authors of Re-frame<sup>[9]</sup> and other Clojure libraries which the Bandish editor application makes extensive use of.

## Bibliography

Courtney, D. n.d. "North Indian musical notation: an overview." Accessed 12 13, 2022. <https://chandrakantha.com/music-and-dance/i-class-music/n-indian-notation/>.

Gandharva, K. 1965. *Anoop Raga Vilas vol one*. Mauj Prakashan Gruh.

n.d. "Indian Oral Tradition." Accessed 12 13, 2022. <https://www.itcsra.org/indian-oral-tradition/>.

Karkera, Kiran. n.d. "Piyarava Bandish in Raga Yaman." *Swaranotebook*. Accessed July 28, 2023. <https://swaranotebook.com/view/2mE9ESZQU7eacLOYsKx3zD9PbV43/8504c37576fd-piyaravav2>.

Ranade, Ashok Damodar. n.d. "Music Contexts: A concise dictionary of Hindustani Music." 73. Bibliophile South Asia Page.

Sawant, Shivraj. n.d. *Vishwamohini Melody Player*. Accessed 07 28, 2023. <http://vishwamohini.com/music/home.php>.

Schooling, Unknown - National Institute of Open. n.d. "Notation system of Hindustani Music." Accessed 07 28, 2023. [https://www.nios.ac.in/media/documents/Hindustani\\_Music\\_242/hindustanimusichtheorybook1/HMB1Ch5.pdf](https://www.nios.ac.in/media/documents/Hindustani_Music_242/hindustanimusichtheorybook1/HMB1Ch5.pdf).

SwarClassical. n.d. *SwarShala* . Accessed 2023. <https://www.swarclassical.com/SwarShala/>.

## Notes

---

1. <https://en.wikipedia.org/wiki/Raga> ↵
2. [https://en.wikipedia.org/wiki/Tala\\_\(music\)](https://en.wikipedia.org/wiki/Tala_(music)) ↵
3. [https://en.wikipedia.org/wiki/Vishnu\\_Narayan\\_Bhatkhande](https://en.wikipedia.org/wiki/Vishnu_Narayan_Bhatkhande) ↵
4. <https://en.wikipedia.org/wiki/Svara> ↵



5. <https://clojure.org/about/spec> ↵
6. <https://github.com/cognitect/transit-clj> ↵
7. <https://github.com/Studio-kalavati/sargam-spec> ↵
8. <https://clojure.org> ↵
9. <https://github.com/day8/re-frame> ↵

# ODD ODDities: some problems with the ODD language unearthed by an attempt to write a processor

Syd Bauman, Northeastern University, United States of America

Helena Bermúdez Sabel, independent scholar

David Maus, State and University Library Hamburg, Germany

Martin Holmes, University of Victoria, Canada

Panel

**Keywords:** ODD, PureODD, ATOP, schemas

## Abstract

Not surprisingly, in the course of attempting to write an ODD processor, the ATOP Task Force has come across several areas where the ODD language is confusingly underspecified. This panel is centered on the ODD language, or more precisely, some of its shortcomings.

## Framing

The session starts with an ~10 minute introduction to the major background concepts that are the infrastructure for the rest of the session. This entails a high-level overview. I.e., mostly circles and arrows, not code. Topics covered include what ODD is, how the Guidelines are built, how customization works, and how customization chaining works.

## Individual Issues

### no duplicate definitions of the same attribute

The ODD language permits an ODD writer to define two different attributes with the same name on the same element (even though XML does not allow this). See <https://github.com/TEIC/TEI/issues/2371>

### Schematron contexts

It is difficult to impossible for an ODD processor to determine the intended context for Schematron assertions that do not have a context explicitly indicated. (In ODD contexts may be specified using the `@context` of `<sch:rule>`, but the `<sch:rule>` is optional.)

### **tagdoc elements allowed most anywhere**

Currently the `model.oddDecl` class is a member of the `model.inter` class, meaning that, e.g., a `<classSpec>` can go inside a `<rdg>`, a `<head>`, or an `<orig>`. Should a processor gather up all tagdoc elements from anywhere in the customization file and assemble them, or should it only process those inside a `<schemaSpec>`? See <https://github.com/TEIC/TEI/issues/2306>.

### **Schematron query language binding**

ODD provides no mechanism for an ODD writer to specify the query language binding. Currently the binding is generated by the software that processes the ODD. We think this is inappropriate: the ODD should specify the query language binding. See <https://github.com/TEIC/TEI/issues/2330>.

### **@generate and @expand of classes**

The `<classSpec>` element has a `@generate` attribute which constrains how the members of the class may be combined; the `<classRef>` element has an `@expand` attribute which asserts how the members of the class should be combined for the given reference. The Guidelines do not:

1. explain what either of these attributes means for an attribute class.
2. define in what order elements should occur when a sequence is generated.
3. explicitly require that the value of `@expand` refers to a pattern that was generated by a `@generate`.
4. explain what should happen when a model class is a member of a model class.

See <https://github.com/TEIC/TEI/issues/2369>.

### **<interleave> (or <bag> or <verschachteln>?)**

To express “all of these elements should be present, but order is unimportant” in a content model, an ODD writer needs to use `<sequence`

`preserveOrder='false'>`. This is, at best, a counter-intuitive way to do this. See <https://github.com/TEIC/TEI/issues/2154>.

### **@require VS @except**

The `<anyElement>` element has two attributes for controlling which elements may be represented: `@require` and `@except`. Either or both may be used; either may have one or more namespace URIs in its value (`@except` also allows QNames of elements). But the Guidelines make no mention of what a processor is to do if the same namespace URI is on both `@except` and `@require`. See <https://github.com/TEIC/TEI/issues/2369>.

## About the authors

**Syd Bauman** is chair of the ATOP Task Force. He served as the N. American Editor of the TEI from 2001 to 2008, during which time the current ODD language was developed; thus he takes partial responsibility for some of the problems discussed herein. He has served as a member of the TEI Council since 2013. ORCID: 0000-0003-3288-443X Affiliation: Northeastern University Digital Scholarship Group

**Helena Bermúdez Sabel** is a Digital Humanities researcher and a software developer at JinnTec. She has been the technical lead of several corpus linguistics and digital editing projects. She has served as a member of the TEI Technical Council since 2021. ORCID: 0000-0002-8627-1367 Affiliation: JinnTec

**David Maus** works as head of research and development at the State and University Library Hamburg. He acts as liason to digital humanities projects at Hamburg University and other higher education institutions. He is author of SchXslt, a modern implementation of ISO Schematron. ORCID: 0000-0001-9292-5673 Affiliation: State and University Library Hamburg

**Martin Holmes** is a programmer/consultant at the University of Victoria Humanities Computing and Media Centre. He is the lead programmer on several large digital edition projects including the Map of Early Modern London, and has served on the TEI Council and as Managing Editor of *jTEI*. ORCID: 0000-0002-3944-111 Affiliation: University of Victoria

# On the Relationship Between Interfaces and TEI in Digital Scholarly Editions

Yannik Werner Herbst, Universität Würzburg, Germany

Torsten Roeder, Universität Würzburg, Germany

Seyed Mohammad Sajad Saderi, Universität Würzburg, Germany

Long Paper

**Keywords:** Interfaces, Digital Editions

## Abstract

Graphical user interfaces, as communication channels between man and machine, represent a crucial ingress route for the use of digital editions. They are an expression of culture-specific forms of access to digital resources, they can be seen as an intentionally designed usage guideline (cf. Dillen 2018), and further as »arguments« of editions themselves (cf. Andrews/van Zundert 2018). Under the paradigm of separating the representational (data) layer from the presentational (output) layer, they serve as a bridge by mediating one into the other, appearing as integral parts of the presentation determined by the underlying data.

The steep increase in the number of digital edition projects has led to repeated ventures to generalize this mediating function and integrate it into frameworks for digital editions. Multiple-use interfaces seem attractive, however their development is costly, and so far they could only be realized convincingly on a relatively basal level or under the condition of sufficiently large similarities of the underlying data (cf. Galka/Klug 2023). An outstanding approach is, for example, the TAPAS project (2016), where the used XSLT script can be selected during use. Fechner (2020) proposed to use a manifest to fill the gap between edition data and their interfaces, based on the example of IIF. Relevant parts of this approach have already been implemented in the ediarum.WEB (2020) software library, and allow digital edition projects to focus development work on data modes and interfaces.

Even though interfaces of digital editions come in a variety of forms, the object description already captured in ediarum.WEB can be translated into reusable components. We will demonstrate how the framework »Synopticon« (cf. Herbst et al. 2023), developed at the Center for Philology and Digitality (ZPD) of the University of Würzburg, takes up this idea and tries to achieve a generic, modular,

and adaptable solution for recurring structures in synoptic interfaces of digital editions. For this purpose, the single TEI-encoded text is modeled as a triad of navigation, text layers, and meta-information in order to juxtapose several such units. The fact that these components are already implicitly contained in TEI documents, queries or document transformations (cf. figure 1), shows how strongly intertwined the interface and the TEI document are, despite the paradigmatic separation of data and presentation layers. From this perspective, the interface appears as a purpose-built derivative of the data layer.

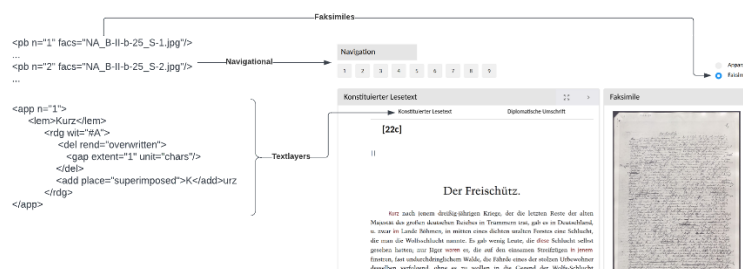


Fig. 1: From TEI to Interface (by Yannik Herbst)

Additionally, to these tasks, the long-term implications of an interface description model need to be discussed. Due to the increasing relevance of digital heritage, the requirement to describe digital interface components in TEI will emerge in the foreseeable future. Digital interfaces constitute para texts (cf. Dillen 2018) which are comparable in their purposes to material text. For these, the TEI provides sufficient elements and attributes. Can an interface description model also provide a means of describing older digital interfaces and, at the same time, support the generic generation of new interfaces? Or are these two different sides of the same coin?

## Bibliography

Andrews, Tara L. and van Zundert, Joris J. (2018): »What Are You Trying to Say? The Interface as an Integral Part of Argument«, in: Digital Scholarly Editions as Interfaces, Schriften des Instituts für Dokumentologie und Editorik, Band 12, hrsg. von Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber und Gerlinde Schneider, Norderstedt: BoD, S. 3–33, <http://nbn-resolving.de/urn:nbn:de:hbz:38-91064>.

Dillen, Wout (2018): »The Editor in the Interface: Guiding the User through Texts and Images«, in: Digital Scholarly Editions as Interfaces, Schriften des Instituts für

Dokumentologie und Editorik, Band 12, hrsg. von Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber und Gerlinde Schneider, Norderstedt: BoD, S. 35–59, <http://nbn-resolving.de/urn:nbn:de:hbz:38-91116>.

ediarum.WEB (2020): Berlin-Brandenburg Academy of Sciences and Humanities, <https://github.com/ediarum/ediarum.WEB>.

Fechner, Martin (2020): A Standardized Interface for Digital Scholarly Editions, Berlin-Brandenburg Academy of Sciences and Humanities, <https://nbn-resolving.org/urn:nbn:de:kobv:b4-opus4-33277>.

Galka, Selina, & Klug, Helmut W. (2023): »Minimal Editing: Die Hyperdiplomatische Transkriptionsplattform«, 9th DHd conference, <https://doi.org/10.5281/zenodo.7715321>.

Herbst, Yannik; Roeder, Torsten & Reul, Christian (2023): »Synoptische Interfaces Digitaler Editionen«, 9th DHd conference, <https://doi.org/10.5281/zenodo.7715462>.

TAPAS (2016): Northeastern University Library's Digital Scholarship Group, <https://www.tapasproject.org/>.

# PDB18 – The German Letter in the 18th Century

**Elisabeth Décultot**, Interdisziplinäres Zentrum für die Erforschung der Europäischen Aufklärung

**Stefan Dumont**, Berlin-Brandenburg Academy of Sciences and Humanities, Germany

**Katrin Fischer**, Universitäts- und Landesbibliothek Darmstadt

**Dario Kampkaspar**, Universitäts- und Landesbibliothek Darmstadt

**Jana Kittelmann**, Interdisziplinäres Zentrum für die Erforschung der Europäischen Aufklärung

**Ruth Sander**, Berlin-Brandenburg Academy of Sciences and Humanities, Germany

**Thomas Stäcker**, Universitäts- und Landesbibliothek Darmstadt

## Poster

**Keywords:** correspondence, letters, CMIF, enlightenment, *correspSearch*

The project establishes a co-operative network and creates a representative digital collection of already edited and printed 18th-century letters. As an infrastructure project in the DFG funding program "Digitisation and Indexing", PDB18 does not itself pursue a research question of its own, but will enable research on epistolary exchange in the 18th century using digital methods (network analysis, text mining etc.) by providing comprehensive data sets of metadata and full texts.

The corpus will include letters that correspondents have written in the German-speaking area or sent there. It potentially includes all languages, although German, Latin and French will predominate. The letters will be accessible via metadata, images, and (in part) full texts via the portal PDB18 ([www.pdb18.de](http://www.pdb18.de)), which is based mainly on the existing infrastructures *correspSearch* (Dumont, Grabsch, Müller-Laackman 2022) and *ZEiD*<sup>[1]</sup>.

The metadata of the edited letters is captured and provided according to the Correspondence Metadata Interchange Format (Stadler 2014, TEI Correspondence SIG 2018). The capturing of metadata and the enrichment with IDs from authority files (Stadler 2012) is done manually, as this has proven in tests to be the most efficient approach. The metadata is aggregated in *correspSearch*, where it can be searched extensively (Dumont 2016) and retrieved via various APIs (TEI-XML, TEI-



JSON; in the future also as RDF). Within the project, correspSearch will be extended by various functions, e.g. new search functions (such as the language used in the letter) and a registry in which letters can be defined as (abstract) work entities. This will also make it possible to map different editions of the same letter to each other.

The published letters from 18th century prints and modern editions are scanned in the ULB Darmstadt. The images form the basis for generating full texts using the OCR software Transkribus. Transformation tools are used to convert the text into a TEI-XML format, which is based on the Base Format of the German Text Archive (BBAW 2022). The TEI-XML documents are stored in eXist-db (eXist Solutions 2023), the transcribed text and the corresponding facsimile are presented side by side with the help of the framework wdbplus (Kampkaspar 2018). All full texts are searchable and offered for download via interfaces.

The project is funded by the Deutsche Forschungsgemeinschaft (DFG) and started in 2022.

## Bibliography

Berlin-Brandenburgische Akademie der Wissenschaften (Ed.). 2022. *DTABf. Deutsches Textarchiv – Basisformat*.

<https://www.deutschestextarchiv.de/doku/basisformat/brief.html>

Dumont, Stefan. 2016. „correspSearch – Connecting Scholarly Editions of Letters“. *Journal of the Text Encoding Initiative* (10). <https://doi.org/10.4000/jtei.1742>

Dumont, Stefan; Grabsch, Sascha; Müller-Laackman, Jonas (Eds.). 2022. *correspSearch – Connect Scholarly Editions of Correspondence (2.1.0)* [Webservice]. Berlin-Brandenburg Academy of Sciences and Humanities 2022.

<https://correspSearch.net>

eXist Solutions. 2022. *eXist-db* [Software]. <https://exist-db.org>

Kampkaspar, Dario. 2018. *W. Digitale Bibliothek (wdbplus)*. [Software].

<https://github.com/dariok/wdbplus>

Stadler, Peter. 2012. „Normdateien in der Edition“. *editio* 26: 174–83.

Stadler, Peter. 2014. 'Interoperabilität von Digitalen Briefeditionen'. In *Fontanes Briefe Ediert*, edited by Hanna Delf von Wolzhagen, 278–87. Fontaneana 12.

Würzburg: Königshausen & Neumann.

TEI Correspondence SIG. 2018. *Correspondence Metadata Interchange Format* (CMIF). <https://github.com/TEI-Correspondence-SIG/CMIF>

## About the authors

**Elisabeth Décultot** has been Professor of Modern German Literary Studies at Martin Luther University Halle since 2015 and has led the Interdisciplinary Centre for European Enlightenment Studies (IZEA) since 2020. She is co-leader of the project PDB18.

**Stefan Dumont** has been a researcher at TELOTA, the DH working group at BBAW, since 2011. He is coordinator of the DFG project "correspSearch" and co-leader of PDB18. His research focuses on digital editions of correspondence and their standardization and networking.

**Katrin Fischer** currently works on PDB18 at the Centre for Digital Editions (CEiD) at University and State Library (ULB) Darmstadt. She holds a degree in History and Applied Geosciences. Her research interests include indexes in scholarly editions, correspondences and mapping historical networks.

**Dario Kampkaspar** is the head of the Centre for Digital Editions in Darmstadt. He holds a *Magister Artium* in history and English philology and has been working on TEI-based projects at the Duke Augustus Library in Wolfenbüttel, the Austrian Academy of Sciences in Vienna and the CEiD for more than 10 years.

**Jana Kittelmann** has been a researcher at the IZEA since 2015. Through her many years of work on various letter editions (print and digital), one of her main research interests is the culture and history of the letter. She is project-coordinator of PDB18.

**Ruth Sander** has been working at TELOTA, the DH initiative of the Berlin-Brandenburg Academy of Sciences and Humanities, since 2021. Her work and research focuses on digital scholarly editions. Since 2022 she is part of the correspSearch team.

**Thomas Stäcker** has been Director of the University and State Library Darmstadt since 2017 and part-time Professor of Digital Humanities at the University of Applied Sciences Potsdam. He is co-leader of the project PDB18. ORCID:

## Notes

---

1. [https://www.ulb.tu-darmstadt.de/die\\_bibliothek/einrichtungen/zeit/index.de.jsp](https://www.ulb.tu-darmstadt.de/die_bibliothek/einrichtungen/zeit/index.de.jsp) ↩

# Pedagogical Approaches to Encoding

Anna Kijas, Tufts University, United States of America

Timothy Duguid, University of Glasgow

Jessica Grimmer, University of Maryland, United States of America

**Workshop**

**Keywords:** digital pedagogy, music encoding,  
text encoding

## Abstract

As practitioners and scholars who engage with music encoding we already understand the affordances that encoding music or text can bring to our research and scholarship. However, individuals who are not yet familiar with encoding schemas and the usefulness of digital methodologies may not have a clear understanding of the “why” in using a digital method or an encoding schema (e.g. MEI, TEI). The aim of this workshop is to provide attendees with a concrete understanding and means for communicating the “why” through the use of two pedagogical approaches: deconstruction/reconstruction and backwards design.

Deconstruction/reconstruction is an approach in which you take an existing project and disassemble it according to the tools, technology, methods, and techniques used in its creation. This approach enables a learner to feel empowered and get an inside or “under-the-hood” look at how a digital project was developed, which in turn allows them to have a better understanding of the affordances that certain digital tools and methods can provide them. Backwards design focuses on identifying learning outcomes at the outset and facilitates learning, as well as critical thinking.

As our first example, we will turn to the Digital Splitleaf project (<https://splitleaf.org>). This project was born out of musical practices such as balladry, folk music, and hymn and psalm singing in which a poetic text is sung with melody with which the performer is most familiar. This practice was particularly common in the early modern period, and it continues today in many churches located throughout Scotland. The printed resources supporting this tradition are produced in a split-leaf format, in which a booklet of tunes is bound above a booklet of texts, thereby allowing the user to select a tune and then select an accompanying text. The result was still rather awkward, requiring singers to keep one eye on the melody

above and another on the text below. Before the creation of the Digital Splitleaf, digital editions would either print the text without the melodies or would preserve the awkward layout of the printed editions with the melody reproduced above and the text below. We will walk through the process of surveying existing digital editions that predated the Splitleaf and then ask workshop participants to consider how they might build their own edition. We will then consider the technologies that were eventually employed with the Digital Splitleaf, offering opportunity for extension and critique. We will then provide a sampling of other projects that use MEI and/or TEI, and through facilitation and prompts attendees will have an opportunity to deconstruct a project with a partner.

Using the backwards design approach, we will ask attendees to identify a learning goal, either for themselves or for students, in which an encoding language/schema plays a role. Attendees will work with a partner to develop questions and learning outcomes specific to their use case. For example: if someone is teaching a music seminar in which they are exploring critical editions in music, they may want to create an assignment in which their students have to encode excerpts from a manuscript or early edition. They will also create a learning activity that they can take away with them and apply in a real-world scenario.

This workshop is for individuals who are interested in applying pedagogical approaches to the encoding of musical and/or textual content. Participants will need their own laptops.

## About the authors

**Dr. Tim Duguid** is a lecturer in Digital Humanities and Information Studies at the University of Glasgow. His current research interests lie in the intersection between digital humanities and historical musicology. His work on Reformation history and early modern music resulted in the creation of a performing edition of the *Wode Psalter*, an early modern music collection, and he was the associate editor for the digital project, "[Letters in Exile: Documents from the Marian Exile.](#)"

**Dr. Jessica Grimmer** is a Registration Specialist in the U.S. Copyright Office and Project Archivist at University of Maryland's Special Collections in Performing Arts. Her current music research focuses on the musical resonances of the Holocaust, both in Europe and in North America. Grimmer is a co-researcher with University of Maryland iSchool assistant professor Katrina Fenlon on the [Sustaining Digital](#)

[Community Collections project](#), which received funding from the Andrew W. Mellon Foundation.

**Anna Kijas** is Head of Lilly Music Library at Tufts University. She is interested in the exploration and application of digital humanities tools and methods in historical (music) research, and in the application of standards, including TEI and MEI, for open access research and publishing, and the use of minimal computing. She leads [Rebalancing the Music Canon](#), an open-access music data project that aims to make works by un(der)-represented people more discoverable, decenter the musical canon, and make data-driven music scholarship more diverse and inclusive.

# Publishing TEI and MEI with TEI Publisher

Wolfgang Meier, e-editiones.org

Magdalena Turska, e-editiones.org

Lars Windauer, e-editiones.org

Workshop

Keywords: TEI, MEI, Publishing, ODD

## Abstract

The motivation behind [TEI Publisher](#) is to provide a tool which enables scholars and editors to publish their materials without becoming programmers. This full day workshop will demonstrate how to create a digital edition with MEI and TEI sources using TEI Publisher. TEI Publisher, now in continuous development for more than 7 years, is already used by a number of large and known edition projects as well as many smaller ones.

TEI Publisher is an open source [community effort](#). At its core it implements the [TEI Processing Model](#), which is part of the TEI standard, and defines in a descriptive, media-independent way how a TEI document should be rendered for publication.

However, online editions require more than just a text transformation: from navigation, pagination, search, facsimile display to handling of mathematical formulas or embedded music notation, and so on. TEI Publisher deals with those aspects, providing all the necessary lego-like building blocks for an online edition.

The workshop will start with a quick introduction to the TEI Processing Model, explaining the techniques needed for transforming a specific TEI document for display. It is recommended that attendants already familiarize themselves with the Processing Model, e.g. by reading through the online workshop available on [GitHub](#).

We will work on a few correspondence examples, which also include musical notation encoded in MEI. Attendants will learn how to:

- output components dedicated to properly display the music embedded within the text
- change the overall layout via templates
- customize the full text and faceted search

- use the TEI Processing model to render additional metadata contained in the MEI.



# Reconciling corpora from enthusiast and scholarly communities: a sustainable approach to augmenting the Electronic Corpus of Lute Music

Tim Crawford, Goldsmiths University of London, United Kingdom

David Lewis, University of Oxford e-Research Centre

Long Paper

**Keywords:** MEI, Tablature, Lute music, Musicology, Software Sustainability

## Abstract

In this paper we report on the augmentation of the Electronic Corpus of Lute Music (ECOLM) – an online collection of encoded music – with material originating from different communities with somewhat different priorities. These differences are reflected in the selection of music, the style of the encodings themselves, the level of annotation of the music, and the amount and quality of supporting metadata. Since the new material contains metadata and transcriptions for tens of thousands of items, an important consideration is sustainable and scalable approaches to the integration, improvement and management of the resulting dataset.

Music for the western lute was written in various forms of tablature notation. The repertory (c.1500–c.1800) is very large, perhaps up to 100,000 items surviving in both printed and manuscript sources. Although many are duplicated, these are very rarely identical copies. A large number are lute arrangements of well-known vocal music, often with highly virtuosic elaborations added to the sequences of notes in the original voice-parts.

Variation between versions, and the relationship between arrangements and their models, can provide musicologists with evidence for studies of style, influence and performance practice. To use the methods of digital humanities we need large datasets of encoded lute music, ideally in parallel with contemporary vocal and keyboard music. Digital editions can also allow the music to be presented in a variety of alternative forms from the same encoding: perhaps in a different style of

tablature, as a staff-notation transcription intelligible to non-players, or as audio playback.

The ECOLM project began in 1999 to work on ways of handling lute music in machine-readable formats taking advantage of current web technology. A major step forward since those early days has been the establishment of an MEI module for encoding tablature. Using *Verovio* in a web browser allows lute tablature to be displayed in its Italian and French versions directly from MEI files, and there exists a software converter between most encoding formats used for lute tablature.

ECOLM currently contains around 2,000 academically-curated items, a small subset of the lute repertory. We are augmenting this corpus greatly by incorporating resources intended for a different community – that of lute-playing enthusiasts, whose priorities in certain respects differ from those of academics.

Enthusiast lutenists generally want to play from ‘performing’ editions of lute works; these may correct errors, regularise barring and repeats, reconstruct missing passages, and even combine sections from different sources. Often these interventions are made without comment. A significant number of players will also happily download and play from facsimiles of the original sources as PDFs, accepting or ignoring the often frequent errors or idiosyncracies of those sources. Some, notably professional performers preparing for concerts or recordings, may be prepared to go to the trouble of comparing a performing edition note-for-note with a facsimile, but many will not.

Some 8,000 pieces forming the contents of various published editions have also been donated to us; these comprise the encodings used to generate printed tablature, plus accompanying material (unstructured metadata) in Word or PDF files. The majority were originally published by the UK Lute Society in newsletters or tablature editions distributed to members over several decades (the machine-readable encodings have often been made or revised by the editor, John Robinson, more recently). The supplied metadata is generally quite complete, although over time (e.g. as manuscripts in private hands have changed owner or found their way into public collections) some inevitable discrepancies have crept in.

Meanwhile, we are also adding a larger collection of about 20,000 pieces from an online web-resource <https://www.lutemusic.org>, curated and largely encoded by Sarge Gerbode but contributed to by a number of other enthusiasts, including many encodings of entire sources. An accompanying spreadsheet contains basic metadata for much of the collection. While the spreadsheet provides useful information about

the sources used by the encoders, it has little scholarly pretension and has not been automatically updated as the resource grows.

Finally, we hope to incorporate a large metadata repository for lute sources <https://mss.slweiss.de> which includes full inventories of around 900 sources, plus encoded tablature incipits for over 60,000 pieces. It covers the historical repertory nearly completely, but is of variable accuracy and consistency. On the other hand, it is the only online resource of its kind, offering something like a parallel for lute tablatures to RISM's indexes of polyphonic vocal music sources from the Renaissance.

# Relation<sup>3</sup>: How to relate text describing relationships with structured encoding of the relationships?

Selina Veronika Galka, University of Graz, Austria

Georg Vogeler, University of Graz, Austria

Short Paper

**Keywords:** assertive edition, personal relationships, TEI, RDF, RDFa

## Abstract

How can personal relationships ("Henry and his brother") in historical sources be modeled under the aspect of "assertive editing"? The concept of "assertive edition" is based on a multi-layered representation by linking embedded annotations (annotations with TEI/XML) with external information structures (e.g. RDF) in order to model the facts contained in the text as assertions (Vogeler 2019, 315), whereby different possibilities for this linking exist (Boot/Koolen 2021).

For modeling relationships of persons, the TEI proposes the element [tei:relation](#) with the attributes @name, @active and @passive. We follow this proposal for the modeling of the relationships themselves using the attributes @type for more precise specification of the relation, e.g. "family", @name for the type of relation, e.g. "brother" and @mutal, @active or @passive for the persons involved ("Henry", "Henry's Brother"). This modeling approach is simple, can be easily translated to RDF and as much information as possible remains in the TEI/XML data. However, the TEI does not declare a canonical way to link prosopographical information with text. For linking TEI/XML and RDF structures, RDFa can be embedded directly in TEI/XML; `tei:xenodata` offers a possibility to embed RDF statements in the TEI document. One argument against the use of both of them is the more difficult human readability of the TEI document (Vogeler 2019, 317). As `tei:xenodata` is defined in the context of the `tei:teiHeader`, its main purpose is in the storage of metadata and, by this, seems to be out of scope.

We suggest annotating the text with `tei:seg` and referring to the `@xml:id` of the `tei:relation` with the attribute `@ana`. `@ana` seems to be the best solution to refer to

statements, as @ref is used to link to an entity (Vogeler 2019, 317) and @ana is a globally available attribute. Schwartz, Gibson and Torabi also dealt with modeling relationships in their projects Syriaca.org/SPEAR, where they create a full blown factoid prosopography with TEI. They rely, similar to our approach, on the use of tei:relation to represent the familiar relationships, but the linking method applied is much less granular, as the general factoid model does not require a detailed annotation of the source on which the claim of the factoid is based.

To link the relations and the text, it would be also possible to use tei:annotation - it is following the web-annotation model pointing into the text (annotation/ptr), a referencing method one could call "inward pointing" from the perspective of editorial sciences, while one could call our method, i.e. linking the <relation> to the text as context to which we point via the @ana mechanism from the edited text, "outward pointing". Both the tei:xenoData and the tei:annotation introduce an additional effort in encoding with not adding more information necessary for our edition: the tei:annotation would add a richer description of responsibility and documentation of the annotation act, while the tei:xenoData approach would add rich linked data semantics which would only help to include external definitions of the relation types, while all other modeling semantics is available in the TEI itself. In the context of our edition this additional effort is not justified. Therefore, we decided to use the most simple encoding pattern. The encoding can be converted via XSLT into both other methods without losing information, see [GitHub: MGS](#). Additionally, the outward pointing method is better aligned with the layered editorial concepts as presented by Elena Pierazzo (2015, 43) or Patrick Sahle (2013:III, 251-340), where the annotation is considered an interpretation of the text passage, i.e. an analytic step based on the text itself.

With these modeling proposals, no new TEI elements or attributes would need to be defined, all information is available in the TEI/XML and can be mapped to RDF with appropriate transformations.

## Bibliography

Boot, Peter and Koolen, Marijn. "Connecting TEI Content Into an Ontology of the Editorial Domain.". In *Graph Data-Models and Semantic Web Technologies in Scholarly Digital Editing*, edited by Elena Spadini, Francesca Tomasi, and Georg Vogeler, 9-29. Norderstedt: BoD, 2021.

Pierazzo, Elena. *Digital Scholarly Editing: Theories, Models and Methods*. Farnham: Ashgate, 2015. <https://journals.openedition.org/variants/pdf/408>.

Project: SAWS. Composed of teams working at [King's College London](#), the [Newman Institute](#), Uppsala, and the [University of Vienna](#). URL: <https://ancientwisdoms.ac.uk/> [2023-04-15].

Project: Syriaca.org. Vanderbilt University, Princeton University, and the Contributor(s), 2014. URL: <https://syriaca.org/index.html> [2023-04-13].

Project: SPEAR. Syriac Persons Events and Relations. <https://spear-prosop.org/index.html> [2023-08-01].

Sahle, Patrick. *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: Books on demand, 2013.

TEI: <relation>. Version 4.6.0. Last updated on 4th April 2023, revision f18deffba. URL: <https://tei-c.org/release/doc/tei-p5-doc/en/html/ref-relation.html> [2023-04-15].

Schwartz, Daniel L., Nathan P. Gibson, und Katayoun Torabi. „Modeling a Born-Digital Factoid Prosopography Using the TEI and Linked Data“. *Journal of the Text Encoding Initiative*, 21. März 2022. <https://doi.org/10.4000/jtei.3979>.

Vogeler, Georg. "The 'assertive edition': On the consequences of digital methods in scholarly editing for historians." *International Journal of Digital Humanities* 1 (2019): 309–322.

## About the authors

**Selina Galka** is a research project assistant at the Institute Centre for Information Modelling (University of Graz), currently working in the project *Digital Edition of the Memoirs of Countess Schwerin (1684–1732)*. Her main research interests are digital edition and data modeling.

**Georg Vogeler** is professor for Digital Humanities at the Institute "Centre for Information Modelling" at the University of Graz. He served in the TEI board of directors in 2018–2019. In the Digital Humanities his research interest lies in Digital

Scholarly Editing, Semantic Web technologies, Data Modelling, and application of Data Science to the Humanities.

# Selective Encoding: Reducing the Burden of Transcription for Digital Musicologists

**Mark Saccomano**, Paderborn University, Center for Music, Edition, Media  
(ZenMEM)

**Lisa Rosendahl**, Paderborn University, Center for Music, Edition, Media (ZenMEM)

**David Lewis**, University of Oxford e-Research Centre

**Andrew Hankinson**, RISM Digital Center, Switzerland

**Johannes Kepper**, Beethovens Werkstatt, Germany

**Kevin Page**, University of Oxford e-Research Centre

**Elisabete Shibata**, Beethoven-Haus Bonn, Research Centre "Beethoven-Archiv"

**Long Paper**

**Keywords:** Music Encoding, Metadata,  
Arrangements, Data Modeling,  
Corpora

## Abstract

One of the largest barriers to digital musicology is the time required to create an encoded music file. These files contain the musical data that often serve as the primary material in digital research. While tools exist to automate parts of the process, (format converters for music notation software, for example, and optical music recognition technology) most of the symbolic content—pitches and rhythms—still needs to be entered manually, note by note. Even for an experienced encoder, it can take up to two weeks to finish encoding a single symphonic movement.

To facilitate the creation of corpora for digital analysis, we have developed a procedure for encoding only the portions of a score relevant to a particular study. These encodings can then be extended at a later time, and by any other scholars who have access to them. Currently, there is no standard way to record metadata that details which specific aspects or sections of a score have been encoded. This paper will introduce a pair of possible methods, constructed in the course of our research and tool development, to enhance the ability of MEI to accommodate these partial encodings and then evaluate the benefits of using one approach over the other.



The first method takes advantage of MEI's capacity to create customized schemas. It features both new and repurposed elements and attributes to model a music encoding as the complete collection of a source's encoded and unencoded sections, with each section having a unique identifier and a set of parameter values that allow for easy-to-find and easy-to-read metadata regarding the extent of the score's digital representation.

The second, simpler method works within current MEI structures and consists of additional documentation to clarify existing usage, taking advantage of element entailments along with text descriptions to identify and retrieve encoding information from an MEI file.

We present our experience in creating the corpus for the Beethoven in the House project, a digital study of Beethoven's symphonic works arranged for performance in the home. Our project serves as a case study that illustrates some of the assumptions that underlie these two methodologies, and how project-based considerations can lead to the adoption of one approach over another. Time and resources were limited, and encoding entire symphonic scores was neither feasible nor necessary, as musicologists on the project were only interested in comparing certain passages of the large scale works and their realization in various arrangements. This led to our interest in developing and documenting methods in MEI for creating what we call "selective encodings."

We discovered that the introduction of new elements and attributes in MEI can be at odds with an archival philosophy that prioritizes the preservation of materials and interoperability of digital resources. And while a project may have needs that can be easily addressed with additional data structures, it can be worthwhile to consider instead adapting to the data model of an existing standard, thus better ensuring that a project's research contributions can be shared and its data reused.

## About the authors

**Andrew Hankinson** is a researcher and software developer for the RISM Digital Center in Bern. He has held positions on the technical group and board of the Music Encoding Initiative.

**Johannes Kepper** studied music and media science as well as computer science at the Musicology Seminar Detmold/Paderborn and the University of Paderborn.

Since 2006 he has been active in the development of the Music Encoding Initiative (MEI) and is the German PI of the Beethoven in the House project.

**David Lewis** trained as a historical musicologist at Kings College, London. He is currently a Researcher at the University of Oxford e-Research Centre and lecturer in Computer Science at Goldsmiths.

**Kevin Page** is an associate faculty member and senior researcher at the University of Oxford e-Research Centre. He is co-founder of the Digital Libraries for Musicology conference, teaches digital musicology and linked data methods for the Master's programme in Digital Scholarship at Oxford, and is the UK PI of the Beethoven in the House project.

**Lisa Rosendahl** is a research associate on the project Beethovens Werkstatt at Beethoven-Haus Bonn. With master's degrees in history and musicology, as well as a certificate in digital humanities, she brings an interdisciplinary approach to her research on music and social history of the 18th and 19th centuries.

**Mark Saccomano** is a music theorist at Paderborn University and a post-doctoral researcher for the Beethoven in the House project. He previously taught music history and music theory at Columbia University in New York and was adjunct professor of music at Montclair University in New Jersey.

**Elisabete Shibata** focuses on the connection between music and new technologies. She is currently pursuing her PhD under Prof. Dr. Frank Hentschel at the University of Cologne, where she is investigating the digital representation of arrangements using Beethoven's music as an example.

# Studying Poetry through Music: The Tasso in Music Project

Emiliano Ricciardi, University of Massachusetts Amherst, United States of America  
Craig Sapp, Packard Humanities Institute/CCARH/Stanford University

Long Paper

**Keywords:** TEI, Literary Variants, Analysis,  
Music/Text Relations, Tasso

## Abstract

The Tasso in Music Project (<https://tassomusic.org>) is a complete digital edition of the early modern musical settings of the poetry of Torquato Tasso (1544–95), the most prominent poet of late sixteenth-century Italy. Comprising about 800 musical settings and representing the work of over 200 composers, this repertoire is significant not only for musical reasons—for instance, with many parallel settings of the same poems, the corpus lends itself especially well to comparative analysis—but also, and most importantly in this presentation, for literary reasons, since the musical settings can shed light on the dissemination of Tasso’s work and can offer insight into the form and meaning of his poems.

Accordingly, besides providing newly made critical editions of the musical settings in a variety of electronic formats (Humdrum, MEI, MusicXML), the project features a substantial literary component whose function is to help us to better understand the tradition of the poetic texts and music/text relations. More specifically, the project includes TEI transcriptions of the poetic texts as they appear in musical settings and in contemporaneous literary sources, both manuscript and printed, as well as tools for the dynamic visualization of literary variants across sources. This textual component provides indispensable data for the study of the transmission of Tasso’s poetry, which is notoriously intricate as many of his poems survive in multiple versions with substantial variants. These data are of interest for musicologists, who can use them to trace the sources used by composers, to assess filiation between settings of the same poem, and to study possible manipulations of literary texts by composers. The data are equally interesting for literary philologists, who can use them to broaden their perspective on the tradition of Tasso’s texts, having access for first time to variants that are recorded only in musical sources,

some of which may be ascribable to Tasso himself by virtue of his proximity to several major composers of the time.

Likewise, building on the possibilities afforded by digital encoding, the Tasso in Music Project offers several online tools to study the interaction of poetry and music, both within single pieces and across the repertoire. The data for both music and poetry are also available for further offline and additional analyses. Currently available online analysis tools address the melismatic treatment of words and the use of textual repetition in the music—the latter is particularly useful to quickly identify keywords in a poem. In the presentation, we will also present new tools that will allow users to analyze the relationship between poetic prosody and musical durations/meter as well as between poetic and musical syntax (for instance, the correspondence, or lack thereof, between ends of poetic lines and cadences), showing how musical settings can function as parsings of a poem's metric and syntactic structure.

Through these tools, the Tasso in Music Project restores the centrality of poetry in early modern vocal music, addressing an interdisciplinary audience encompassing not only performers and scholars of music (historians, theorists, music encoders), but also scholars of literature (Italianists, linguists, textual encoders).

# Teaching Encoding In and Out of the Classroom

**Diane Jakacki**, Bucknell University, United States of America

**Brian Croxall**, Brigham Young University, United States of America

**Janelle Jenstad**, University of Victoria, Canada

**Constance Crompton**, University of Ottawa

**Gimena del rio Riande**, University of Buenos Aires

**Emmanuel Nguê Um**, University of Yaoundé 1

**James Cummings**, Newcastle University

**Timothy Duguid**, University of Glasgow

**Kiyonori Nagasaki**, International Institute for Digital Humanities

**Martina Scholger**, University of Graz

**Raffaele Viglianti**, University of Maryland

**Panel**

**Keywords:** pedagogy, learning, minimal computing, multilingualism

## Abstract

TEI is one of the founding technologies of the digital humanities, and text-encoding offers a productive way for students and colleagues to begin engaging with DH. And yet, there is relatively little discussion about TEI (and MEI) pedagogy at this conference, the DH conference, and in the secondary literature. We believe that one reason for this absence is that the teaching of TEI and MEI happens in different environments—undergraduate classrooms, graduate seminars, the training of research assistants, and workshops of varying lengths, intensities, and audiences—to different ends—literary analysis, text recovery, data mining, translation studies, digital editorial production—and with radically different kinds of texts. Increasingly, students are learning TEI to undertake work on cultural reclamation/representation and social justice, for example.

To address this lacuna, we have begun work on a volume that foregrounds the pedagogy of encoding, and this roundtable is a first effort of some of the volume's contributors to present their arguments and to hear feedback from the community. The roundtable panel we propose we hope will address some of these different approaches to teaching and learning TEI and MEI through a series of short talks,

rapid-fire questions from the presiders to the panelists, and a robust 25-minute conversation that involves the entire audience.

Nine panelists will each address teaching TEI in different contexts. Constance Crompton will describe how her undergraduates encode Victorian classics with an eye toward contemporary problems of textual preservation and representation. The students Gimena del Rio Rande teaches often have radically disparate access to computing technologies, which has led her to use minimal computing infrastructures to create equitable digital editions. Relatedly, Raffaele Viglianti will consider what TEI learners must be taught in order to minimize the role of infrastructure in TEI and MEI publishing, with a focus on helping those learners thoughtfully disseminate their digital publications. Working in Africa, Emmanuel Ngué Um's students experience multilingualism as a fact of social life but also confront linguistic uniformization; he will address how text encoding and specifically TEI provide a framework for accommodating linguistic diversity and preserving linguistic rights as a means for social justice. Kiyonori Nagasaki also confronts language challenges in TEI pedagogy, focusing on the challenge of adapting TEI for, generally, non-Latin scripts and, specifically, Japanese. Tim Duguid will contemplate the challenges of teaching MEI simultaneously to media students (who have some technical skills but no musical skills) and music students (who have music skills but no technical skills), including both the work of encoding and analysis. Rather than teaching two types of students, Janelle Jenstad discusses what her students gain by learning two XML languages: TEI and DocBook; comparing the two languages not only helps students understand the flexibility of XML languages but also helps them approach TEI critically as one language among many, with its own particular applications, strengths, and weaknesses. Martina Scholger, who frequently teaches workshops for early-stage researchers at the Institute for Documentology and Scholarly Editing, will reflect on best practices for both pedagogy and curricula development while also examining how the social and community-building aspects of such workshops are critical components of their pedagogy. Finally, James Cummings will present a provocation and argue that it is time to deemphasize XML as the centerpiece of encoding education and to focus from the start on the TEI conceptual model itself.

Our plan for the session is as follows. First, each presenter will speak for three to five minutes, outlining the key argument of their chapter for the volume. Second, the presiders (Jakacki and Croxall) will lead a lightning round in which a question will be posed and each panelist will have a maximum of 60 seconds to respond. The lightning round will have a maximum of two questions, which will be shared with

panelists ahead of time so they have a chance to formulate responses. Our potential questions include the following:

- What was the thing you had to teach yourself in order to teach the TEI?
- How has teaching changed your relationship to the TEI/MEC?
- What outcomes/skills do your students gain from learning encoding? How does encoding change their lives/studies? What was the hardest TEI teaching experience you've had?

Third, the remainder of the time will be devoted to framed discussion among the panelists and the audience.

## About the authors

**Constance Crompton** is a Canada Research Chair in DH and director of the University of Ottawa's Humanities Data Lab. She is a member of the Lesbian and Gay Liberation in Canada, Linked Infrastructure for Networked Cultural Scholarship, and Implementing New Knowledge Environments research teams.

**Brian Croxall** is assistant research professor of digital humanities at Brigham Young University. He is the co-editor (with Diane Jakacki) of *What We Teach When We Teach DH: Digital Humanities in the Classroom* and the co-editor (with Rachel A. Bowser) of *Like Clockwork: Steampunk Pasts, Presents, and Futures*.

**James Cummings** is Senior Lecturer in Late Medieval English Literature and Digital Humanities at Newcastle University, a member of the TEI Board and formerly TEI Technical Council. He directed *Digital Medievalist* (2009-12) and founded *DHOxSS* and *DH Awards*. He has taught many TEI workshops.

**Gimena del Rio Riande** is Associate Researcher at the Instituto de Investigaciones Bibliográficas y Crítica Textual (Argentina) and Professor at the Universidad del Salvador. She is director of the Laboratorio de Humanidades Digitales HD LAB (CONICET) and a postgraduate diploma in digital humanities.

**Timothy Duguid** is lecturer in Information Studies at the University of Glasgow, whose research and teaching interests lie in the intersection between digital humanities and historical musicology. His Digital Splitleaf combines TEI and MEI into a digital edition of early modern metrical psalms.

**Diane Jakacki** is digital scholarship coordinator and associate faculty at Bucknell University. She is the co-editor (with Brian Croxall) of *What We Teach When We*

*Teach DH: Digital Humanities in the Classroom* (U of Minnesota P). She is a Fulbright Canada Research Chair in Digital Humanities, 2022-3.

**Janelle Jenstad** is Professor of English at the University of Victoria. She directs *The Map of Early Modern London* (MoEML) and *Linked Early Modern Drama Online* (LEMDO) With Jennifer Roberts-Smith and Mark Kaethler, she co-edited *Shakespeare's Language in Digital Media* ([Routledge](#)).

**Kiyonori Nagasaki**, a senior fellow at the International Institute for Digital Humanities in Tokyo, has published several books and papers on DH in Japanese. He teaches DH at various universities and recently edited *A Companion to Text Encoding for the Humanities* in Japanese.

**Emmanuel Ngué Um** is Associate Professor of African Languages and Linguistics at the University of Yaoundé 1 (Cameroon). He is currently an Associate Editor for the International Journal of Humanities and Arts Computing (IJHAC), and member of the Editorial Board of the *Journal of Digital History*.

**Martina Scholger** is a senior scientist at the Centre for Information Modelling – Austrian Centre for Digital Humanities at the University of Graz. She has been a member of the Institute for Documentology and Scholarly Editing since 2014 and a member of the TEI Technical Council since 2016.

**Raffaele (Raff) Viglianti** is a Senior Research Software Developer at the Maryland Institute for Technology in the Humanities, University of Maryland. Viglianti is currently an elected member of the Text Encoding Initiative technical council and the Technical Editor of the *Scholarly Editing* journal.



# TEI Lex-0: Recent Developments and New Directions

**Toma Tasovac**, Belgrade Center for Digital Humanities (BCDH); Digital Research Infrastructure for Arts and Humanities (DARIAH)

**Piotr Bański**, Institute for German Language (IDS)

**Axel Herold**, Berlin-Brandenburg Academy of Sciences (BBAW)

**Boris Lehečka**, The Moravian Library in Brno (MZK)

**Laurent Romary**, National Institute for Research in Digital Science and Technology (INRIA)

**Ana Salgado**, The Lisbon Academy of Sciences (ACL); The Linguistics Research Centre of NOVA University Lisbon (CLUNL)

**Panel**

**Keywords:** TEI Lex-0, lexicography, linguistics, dictionaries, grammars, tools

## Abstract

TEI Lex-0 is a strict customization of the TEI P5 Guidelines for marking up dictionaries which establishes a baseline encoding and a target format to facilitate the interoperability of heterogeneously encoded lexical resources (Tasovac et al. 2018). TEI Lex-0 has received strong community support and uptake: for instance, the format was chosen, together with Ontolex-Lemon (McCrae et al. 2017) as one of the default data models by the European Lexicographic Infrastructure (Tiberius et al., 2022); while the DARIAH Working Group on Lexical Resources, which maintains the TEI Lex-0 Guidelines, received the 2020 Rahtz Prize for TEI Ingenuity. [\[1\]](#)

This panel will report on the recent developments with TEI Lex-0 and discuss possible new directions. We will do this by delivering three papers: one on dictionary metadata, one on our user-friendly tools to help both the editing and publication of TEI Lex-0 dictionaries; and one on the challenges of encoding descriptive grammars. We plan to leave sufficient time for community feedback and interaction with the audience. The panel will be of interest not only to TEI experts in lexicography and linguistics, but also to metadata specialists, tool developers and the wider audience of TEI enthusiasts who are interested in the social and technical aspects of the application of TEI in various domains.

## Encoding Metadata for Dictionaries in TEI Lex-O

This paper highlights the need for domain-specific metadata standards and demonstrates the usefulness of strict TEI Lex-O Guidelines for encoding dictionary metadata. While the FAIR principles of findability, accessibility, interoperability and reusability (see Wilkinson et al. 2016) have received wide political support and are well-established in science, technology and innovation domains (see Tóth-Czifra 2020), their generic and discipline-agnostic nature leaves significant room for improvement. Lexicographers and researchers need specific information that goes beyond generic metadata formats and focuses on the specific questions of linguistic scope, structural model, markup granularity etc. FAIR Data on its own is not sufficient for such queries.

The paper will review the current state of metadata in lexicographic resources, with references to both LexBib and LexMeta (see Lindemann et al. 2018; Kosem and Lindemann 2021; and Lindemann et al. 2022), explain the rationale for the strict encoding choices made in TEI Lex-O, and contextualize the work on standardizing TEI Lex-O metadata in the `teiHeader` within the broader standardization landscape including: a) the ongoing work on ISO 24612 (Lexical Markup Framework) (see Romary et al. 2019); and b) the *Lexicographic Data Seal of Compliance*, a community-based certification system for lexicographic resources that adhere to best scholarly practices (Tasovac et al. 2021).

## Building an Infrastructure for TEI Lex-O

This paper describes two user-friendly tools for working on and with TEI Lex-O encoded dictionaries: a customized TEI Lex-O framework for oXygen XML Editor<sup>[2]</sup>, and a customization of the TEI Publisher<sup>[3]</sup> for lexicographic resources.

The TEI Lex-O framework, available as an add-on for oXygen XML Editor, offers access to frequently used pieces of XML code (e.g. namespaces or basic parts of the entry structure) with associated keyboard shortcuts. The Author Mode offers intuitive interactive elements for editing metadata as well as dictionary content. Rules defined using Schematron with a Quick Fix extension are used for validation, advanced checking and editing. For data analysis (e.g. extracting headword lists, distributions of morphological categories etc.), XSLT transformations are bundled and used in combination with XQuery functions.

TEI Lex-0 Publisher extends the main functionalities of TEI Publisher using XQuery, Webcomponents and JavaScript to offer basic and advanced features for working with monolingual (and, in a future iteration multilingual) dictionaries, including, for instance: browsing, simple and advanced search (combination of multiple parameters), facets, facsimile display, the definition of a REST API for working with dictionary data etc.

## Encoding Descriptive Grammars in TEI Lex-0

Descriptive grammars are one of the cornerstones of the study of language: while dictionaries, broadly speaking, describe the meaning of words, grammar books describe how the words are constructed and how they are put together to form meaningful sentences. And just like dictionaries, grammars created in the past are still of interest to humanities scholars: they document not only the past epochs of particular languages, but also the evolving thought about language as such and its central role in society.

In this paper, we'll analyze the differences between encoding grammars and dictionaries while paying special attention to the interplay of structured elements (ranging from individual morphosyntactically tagged forms to full-blown declensions and conjugations) within the largely narrative text of grammar books. We will argue that the encoding of grammar books can be made semantically more precise and infrastructurally more interoperable through the reuse of structural patterns or "crystals" (Romary and Wegstein, 2012) recommended by TEI Lex-0.

The paper will conclude with suggestions on how the encoding of grammatical information can be improved across the several parts of the TEI Guidelines.

## Bibliography

Kosem, I. and D. Lindemann (2021): 'New developments in Elexifinder, a discovery portal for lexicographic literature'. In: Gavriilidou, Z., L. Mitits and S. Kiosses (eds.): *Lexicography for Inclusion: Proceedings of the 19th EURALEX International Congress, 7-11 September 2021, Alexandroupolis*, Vol. 2. Alexandroupolis, pp. 759-766. <https://euralex2020.gr/proceedings-volume-2>.

Lindemann, D., F. Kliche and U. Heid. (2018): 'LexBib: a corpus and bibliography of metalexicographical publications'. In: *Proceedings of EURALEX 2018. Ljubljana*, pp.

699–712. <http://euralex.org/publications/lexbib-a-corpus-and-bibliography-of-metalexicographical-publications/>

Lindemann, D., P. Labropoulou and C. Klaes. (2022). 'Introducing LexMeta: A Metadata Model for Lexical Resources'. In: *XX EURALEX Conference*, Mannheim, Germany. <https://doi.org/10.5281/zenodo.6897062>

McCrae, J. P., Bosque-Gil, J., Gracia, J., Buitelaar, P. and Cimiano, P. (2017). 'The Ontolex-Lemon model: development and applications'. In: *Proceedings of eLex 2017 conference* (pp. 19-21). <http://john.mccr.ae/papers/mccrae2017ontolex.pdf>

Romary L. and W. Wegstein (2012). 'Consistent modelling of heterogeneous lexical structures'. In: *Journal of the Text Encoding Initiative*, Issue 3 | November 2012. URL: <http://jtei.revues.org/540>; DOI: [10.4000/jtei.540](https://doi.org/10.4000/jtei.540)

Romary L, M. Khemakhem, F. Khan, J. Bowers, N. Calzolari, et al.. LMF Reloaded. AsiaLex 2019: Past, Present and Future, Jun 2019, Istanbul, Turkey. [hal-02118319](https://hal.archives-ouvertes.fr/hal-02118319)

Tasovac, T., Romary, L. et al. (2018). *TEI Lex-0: A baseline encoding for lexicographic data*. Version 0.9.1. DARIAH Working Group on Lexical Resources. <https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html>.

Tasovac, T., Romary, L., Tóth-Czifra, E., Marinski, I. (2021). *Lexicographic Data Seal of Compliance. Research Report*. ELEXIS; DARIAH. ([hal-03344267](https://hal.archives-ouvertes.fr/hal-03344267)).

TEI Consortium. (2022). *TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 4.6.0*. Last updated on 4th April 2023, revision f18deffba. TEI Consortium. <https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html>.

Tiberius, C., S. Krek, M. Mechura, J. McCrae and Toma Tasovac (2022). *D1.5 Best practices for lexicography - final report*. European Lexicographic Infrastructure (ELEXIS). [https://elex.is/wp-content/uploads/ELEXIS\\_D1\\_5\\_Best\\_practices\\_for\\_lexicography.pdf](https://elex.is/wp-content/uploads/ELEXIS_D1_5_Best_practices_for_lexicography.pdf)

Tóth-Czifra, E. (2020). 'The Risk of Losing the Thick Description: Data Management Challenges Faced by the Arts and Humanities in the Evolving FAIR Data Ecosystem'. In: J. Edmond (ed.), *Digital Technology and the Practices of Humanities Research*. Cambridge, UK: Open Book Publishers. <https://doi.org/10.11647/OBP.0192>

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). 'The FAIR Guiding Principles for scientific data management and stewardship'. in: *Sci Data* 3, 160018.

DOI: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)

## Notes

---

1. <https://tei-c.org/2020/11/23/2020-rahtz-prize-for-tei-ingenuity/> ↵
2. [https://www.oxygenxml.com/xml\\_editor.html](https://www.oxygenxml.com/xml_editor.html) ↵
3. <https://teipublisher.com> ↵

# TEI Processing and Computational Analysis on Encoding Judgment Text

Hsuan-Lei Shao, National Taiwan Normal University, Taiwan

Sieh-Chien Huang, National Taiwan University, Taiwan

Short Paper

**Keywords:** Judgment Text, Legal Analytics, TEI Application, Empirical Studies, Machine Learning.

## Abstract

### Research Background

Legal documents and judgments serve as crucial sources of legal knowledge and social background information. Analyzing these texts using natural language processing (NLP) techniques holds significant value. However, the complex structure and unique formats of legal documents pose challenges to such analysis. In light of this, our research proposes a specialized form and process for the Text Encoding Initiative (TEI) applied to legal judgments. This process leverages domain-specific legal knowledge to add specific markup for legal concepts and terminology as metadata.

In this article, we incorporate TEI as an integral part of the research design for legal analytics. Our study utilizes a set of homicide judgments as research materials and devises a process to convert raw text into a computable dataframe. This process involves employing specialized data structures for text encoding, methodologies for encoding via XML platforms, structural editing, meta description, and computational analysis

### Literature Review

In the beginning, the TEI process was primarily used in the research fields of literature, history, and philosophy [1,2]. Its application was primarily focused on text data. However, over time, the scope of TEI has expanded significantly. The TEI process is now being applied to various other types of data, such as religious classics [3,4], parliamentary records [5], and even "Voice Data" [6], which is another topic of discussion at this conference. Despite this broadening range of TEI applications,

there is still a noticeable dearth of TEI implementation and annotation within the legal text domain. It is this void that our research team seeks to address and contribute to.

With TEI technology advances, scholars are increasingly reevaluating the ontologies of their respective research fields [7]. They are striving to redefine metadata and (re)build ontologies to suit the digital methods and tools available. In the realm of legal studies, which has a long-standing tradition of matured methodologies, our contribution lies in creating a new ontology specifically tailored to legal research that effectively integrates digital methods and practices. It is also to enhance TEI Engineering by developing/rethinking an “digital ontology” of legal domain.

## **Research Design**

In this case, our primary research objective is to predict the "sentence term of a homicide case." To achieve this, we have designed a comprehensive "sentencing factors" ontology derived from the criminal law code. Through the application of our TEI processing and computational analysis approach, we effectively extract valuable information from legal judgments and employ it to predict the potential sentencing outcomes.

Our research showcases the efficacy of our TEI method in legal text analysis and TEI processing. The adoption of specialized data structures, encoding methodologies, and meta descriptions ensures accurate computational analysis, leading to meaningful results in the context of sentencing predictions. The utilization of TEI in the legal domain showcases the potential of this approach as a valuable tool for enhancing legal text analysis and supporting decision-making processes.

Furthermore, our research extends beyond the legal domain. By offering insights into specific corpora, formal ontologies, and best practices in TEI, our approach has far-reaching implications for researchers in diverse fields. During the conference, our presentation will not only include a comprehensive description of our TEI processing approach, but we will also provide a demonstration of TEI in action within the realm of legal studies. Overall, our research contributes significantly to the broader exploration of TEI's potential in advancing legal studies and computational analysis.

## Bibliography

[1] Vertan, C., & Reimers, S. (2012). A TEI-based Application for Editing Manuscript Descriptions. *Journal of the Text Encoding Initiative*, (2).

[2] Soualah, M. O., & Hassoun, M. (2012). A TEI P5 Manuscript Description Adaptation for cataloguing digitized Arabic manuscripts. *Journal of the Text Encoding Initiative*, (2).

[3] McAllister, P. (2020). Quotes, Paraphrases, and Allusions: Text Reuse in Sanskrit Commentaries and How to Encode It. *Journal of the Text Encoding Initiative*, (13).

[4] Wittern, C. (2020). Digital Texts in Practice. *Journal of the Text Encoding Initiative*, (13).

[5] Wissik, T. (2021). Encoding Interruptions in Parliamentary Data: From Applause to Interjections and Laughter. *Journal of the Text Encoding Initiative*, (14).

[6] Emsley, I., & Roure, D. D. (2016). "It will discourse most eloquent music": Sonifying Variants of Hamlet. *Journal of the Text Encoding Initiative*, (10).

[7] Bowers, J., & Romary, L. (2016). Deep encoding of etymological information in TEI. *Journal of the Text Encoding Initiative*, (10).

## About the authors

**Hsuan-Lei Shao** (National Taiwan Normal University), <https://orcid.org/0000-0002-7101-5272>

**Sieh-Chien Huang** (National Taiwan University), [schhuang@ntu.edu.tw](mailto:schhuang@ntu.edu.tw)



# TEI semantic in graph-based digital scholarly editions

Andreas Kuczera, University of Applied Science, Gießen, Germany

Long Paper

**Keywords:** Graph-based digital scholarly editions, TEI semantic

## Abstract

Digital editions are now established as state of the art in the humanities and cultural studies. A large number of source types from different eras are now digitally edited. Thanks to their almost unlimited editorial possibilities, digital edition processes have proven superior to pure print publication (for the current discussion, see Driscoll/Pierazzo 2016). The shift from print to digital edition brings with it a wealth of search and evaluation possibilities and also enables the development and differentiation of semantic markup (summarised by Sahle 2013). The Text Encoding Initiative (TEI) has established itself as an authoritative standard in this field. Correspondence data in particular are the focus of numerous internationally networked research projects. One example of digital letter editions is the European cooperation project *Reassembling the Republic of Letters* led by Howard Hotson and Thomas Wallnig, which has developed overarching standards in the modelling and digital analysis of correspondence data (cf. Hotson/Wallnig 2019, <http://republicofletters.stanford.edu/>). At the national level, digital editions of early modern correspondence such as the projects *PROPYLÄEN. Goethes Biographica* (<https://goethe-biographica.de/>), *Jean Paul – Sämtliche Briefe digital* (<https://www.jeanpaul-edition.de/start.html>), *edition humboldt digital* (<https://edition-humboldt.de/>) and the *Carl-Maria-von-Weber-Gesamtausgabe* (<https://weber-gesamtausgabe.de/de/Index>) are examples of active Digital Humanities research in this field. The web service *correspSearch* of the Berlin-Brandenburg Academy of Sciences and Humanities (BBAW) should also be mentioned in particular, which offers the networking of numerous editions and thus enables research perspectives beyond the individual edition.

It should be noted, however, that these editions focus on visual access to the material and hardly provide for data-based access.

Graph-based digital editions offer both visual and data-centred access. They thus enable the analysis of semantic information in networked form and transfer it into knowledge graphs (Kuczera 2020). Thus, not only multiple representations adapted to the respective interface become possible from a structured text (such as HTML, PDF, ePub etc.), but also pattern-based queries on content aspects. In these functions, the graph-based digital edition emphasises the usage aspect over the reading function: 'not merely reading, but using the edition' (Gabler 2010). This usage aspect manifests itself, among other things, in the finely granular search and filter options of the web and data interfaces. The search functions of graph-based digital editions are semantically oriented and can show connections in the strongly networked research data and make lines of connection visible. This gives users more room for their own content analyses, e.g. via faceted searches, via data mining methods or semantic or pattern-based search options (Hörnschemeyer 2017). A graph structure that models data flexibly linkable as nodes and edges of a network is particularly suitable for this, as it can also map overlapping multidimensional annotation hierarchies (Kuczera 2020). This allows different annotation levels, e.g. layout, formatting, semantics, content indexing and annotation, to be modelled and evaluated together. Examples of the implementation of digital graph-based editions are the projects *Die Sozinianischen Briefwechsel* (<https://sozinianer.mni.thm.de/home>) and *Das Buch der Briefe der Hildegard von Bingen. Genesis – Structure – Composition* (<https://liberepistolarum.mni.thm.de/home>). Both rely on graph technologies through the combination of graph database (neo4j) and graph-based edition environment in conjunction with TEI semantics, both in data modelling and in editorial practice (Kuczera 2020).

The current state of digital editions is the use of XML. Here is an example:

```

TEI
1 <?xml version="1.0" encoding="UTF-8"?>
2 <TEI>
3 <text>
4 <body>
5 <div xml:id="Ms_germ_fol_841" next="#Ms_germ_fol_842">
6 <div type="session" n="1">
7 <p><b facs="#f0011" n="7." /> erblicken wir einen großen Unter&#x017F;chied
8 zwi-<lb/> &#x017F;chen den entferntern u. nähern Planeten<lb/><note
9 place="left"><hi rendition="u">Zwei be&#x017F;ondere
10 Planeten-Sy&#x017F;teme</hi><lb/></note>von der Son&#x0303;e. <hi
11 rendition="u">Dies giebt zwei be&#x017F;ondere<lb/> Sy&#x017F;teme</hi>.
12 Die Scheide machen die kleinern<lb/> Körper die &#x017F;ich zwi&#x017F;chen
13 Mars u. Jupiter<lb/> bewegen, die ein ganz eignes Sy&#x017F;tem<lb/> bilden,
14 von denen die Ve&#x017F;ta als die<lb/><hi rendition="u" hand="#pencil"
15 >größte</hi>
16 <choice><sic>ungefähr</sic><corr resp="#CT">ungefähr</corr></choice> die
17 <choice><abbr>Oberfl.</abbr><expan resp="#CT"
18 >Oberfläche</expan></choice> von Deut&#x017F;ch-<lb/> land hat. Sie
19 haben eine translative Be-<lb/> wegung von We&#x017F;ten nach O&#x017F;ten,
20 &#x017F;ind ihrer<lb/> Stellung nach ähnlich den
21 <choice><sic>Com&#x0303;eten</sic><corr resp="#BF"
22 >Cometen</corr></choice>; obgleich<lb/> doch keine A-<subst><del
23 rendition="u">row"><gap reason="illegible" unit="chars" quantity="1"
24 /></del><add place="across">e</add></subst>-hnlichkeit anderweit
25 zwi&#x017F;chen<lb/> ihnen u. den
26 <choice><abbr><choice><sic>Com&#x0303;et&#xFFFC</sic><corr
27 resp="#BF">Comet&#xFFFC</corr></choice></abbr><expan
28 resp="#BF"><choice><sic>Com&#x0303;eten</sic><corr resp="#BF"
29 >Cometen</corr></choice></expan></choice> i&#x017F;t, wie
30 überhaupt<lb/> kein Uebergang zwi&#x017F;chen Planeten u. Co-<lb/> meten
31 gefunden wird u. keine po&#x017F;itive<lb/><note place="left"><hi
32 rendition="u">Er&#x017F;tes Sy&#x017F;tem<lb/> charakteri&#x017F;t.
33 Merkmale</hi><lb/></note>Ähnlichkeit. <hi rendition="u">Jn
34 die&#x017F;em doppelten Sy&#x017F;tem<lb/> der Planeten gehören
35 zu&#x017F;am&#x0303;en: Merkur,<lb/> Venus, Erde, Mars.</hi> Sie haben
36 das<lb/> gemein&#x017F;ame der be&#x017F;ondern Dichtigkeit,<lb/> wie <hi

```

Fig. 1: XML-Page from the German Text Archive

Text and annotations are captured together in XML structures. The text still remains readable in our simple example. However, it is obvious that with more annotations, which may also come from different people, the complexity increases very quickly. In addition, with XML the so-called containment applies, i.e. different annotations may not overlap. A new approach to store text and annotations is the modeling of text and annotations in a graph.

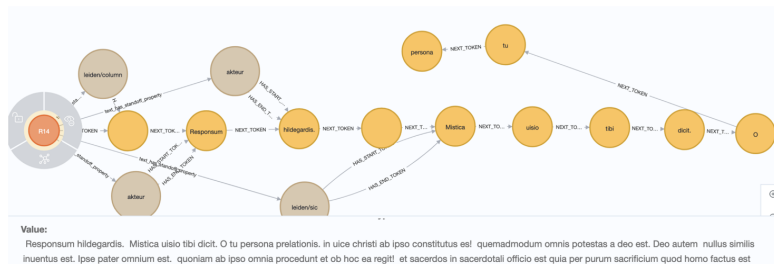


Fig. 2: Text as a chain of Token-Nodes accompanied by annotation nodes which are not part of the Token-Chain

In this case, the words are put into (yellow) token nodes, context. Annotations are attached to this text chain in additional (brown) nodes.

Thomas Efer has also described such a model in his dissertation.

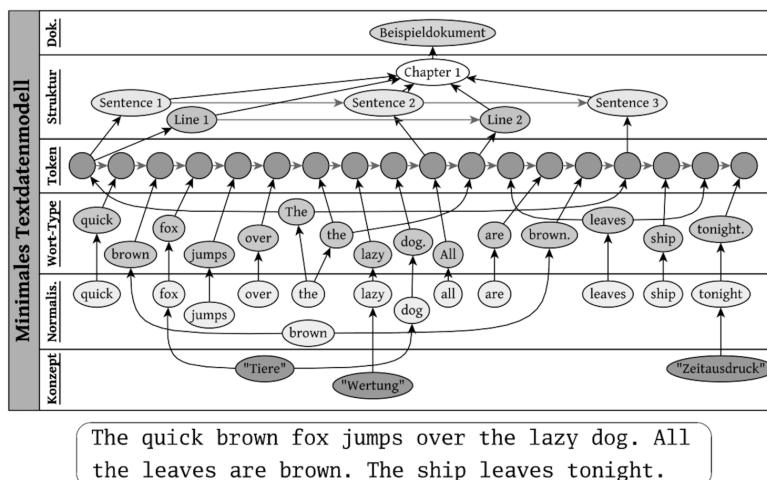


Fig. 3: Graph-based text model from (Efer 2016) p. 76

(Efer 2016) ends at the token level. In our use case, another layer is added with the individual characters in the nodes, as shown in the following figure.

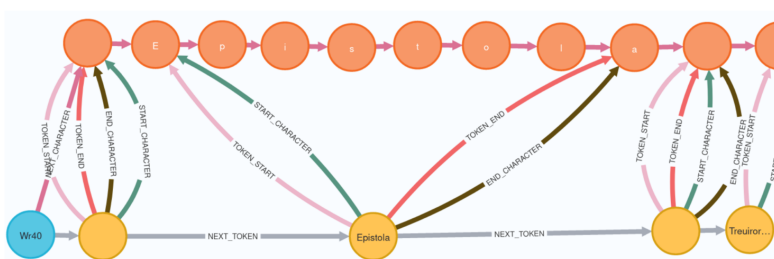


Fig. 4: Text model with token and character level based on texts from the Hildegard project

(see also: <https://liberepistolarum.mni.thm.de>).

The Character-Chain models whitespace as explicit nodes, so the model can be agnostic about what is to be defined as a token.

The presentation will show the graph-based scholarly editions of SBW and Hildegard which aim to give full flexibility within the annotation process to the editor while keeping interoperability with using the semantic of the TEI.

## Bibliography

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

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

Hörnschemeyer 2017: Jörg Hörnschemeyer, *Textgenetische Prozesse in Digitalen Editionen*, Diss. Köln 2017.

Hotson/Wallnig 2019: Howard Hotson, Thomas Wallnig (Hg.), *Reassembling the Republic of Letters in the Digital Age. Standards, Systems, Scholarship*, Göttingen 2019.

Kuczera 2020: Andreas Kuczera, 'TEI Beyond XML – Digital Scholarly Editions as Provenance Knowledge Graphs'. In: Tara Andrews, Franziska Diehr, Thomas Efer, Andreas Kuczera, Joris van Zundert (Hg.), *Graph Technologies in the Humanities – Proceeding 2020*, published at <http://ceur-ws.org/Vol-3110/paper6.pdf>.

Sahle 2013: Patrick Sahle, *Digitale Editionsformen. Zum Umgang mit der Überlieferung unter den Bedingungen des Medienwandels, Textbegriffe und Recodierung*, 3 Bde., Norderstedt 2013 (Schriften des Instituts für Dokumentologie und Editorik, Bde. 7–9).

## About the author

**Andreas Kuczera** is professor for applied digital methodology in the humanities and social sciences at the university of applied science Gießen. He works on graph technologies in the digital humanities and graph based digital scholarly editions.

# TEI XPointer Schemes – Implementation and Example Application

Christian Lück, University of Münster, Germany

Ludger Hiepel, University of Münster, Germany

Johannes Schnocks, University of Münster, Germany

Short Paper

**Keywords:** XPointer, annotation, LOD, stand-off, Bible, referencing fragments

## Abstract

The XPointer framework is one of those W3C notions that are dead today. So are the TEI XPointer schemes. The reasons are manifold. Most notably, there has never been a full-grown implementation of an XPointer processor. Furthermore, it is unclear, what a processor is to do with the pointers, that actually only *point* to a portion of a resource. XPointers have thus never grown to their paradigmatic field of application, in which they could have proven their value. Specified in the SATS section of the guidelines, TEI XPointer schemes occur at several other sections, among them dead-by-design SASO. This bad company appears to be symptomatic.

However, there is a paradigmatic field of application for the TEI XPointer schemes: the Web-Annotations-like `<annotation>` element and its `@target` attribute. If this element is really meant to “represent an annotation following the Web Annotation Data Model” like the TEI reference states, then we need a referencing mechanism, that is compatible with the Web Annotation’s selector mechanism, and that at a specification level. The TEI XPointer schemes not only do satisfy this requirement but are the only specified component of the TEI, that satisfies it.

This short paper first introduces a full implementation of the TEI XPointer schemes: an ANTLR-based XPointer parser, a Saxon-based processor, and its APIs.<sup>[1]</sup> The processor has a Java API and an XPath API. This enables the evaluation of XPointers in XSLT and XQuery through XPath function extensions, e.g., `xptr:get-sequence(...)` for getting the sequence of items pointed to, or `xptr:type(...)` to get the pointer’s type. There are also conversion functions for translating TEI XPointers to Web Annotation Selector schemes and vice versa.

The paper then showcases the application of TEI XPointers in `annotation/@target` in a project from the field of Old Testament research. Here, portions of Masorettes' book of Ijob are annotated and related to translations in Septuagint and Targum, in order to examine semantic deferrals in the textual traditions. This is a specific application in the more general domain of inquiries on intertextuality. A TEI-based approach proves its value, when there's a mix of stable texts and texts still under editing--provided that the data model will allow straightforward transformation to linked open data. The TEI XPointers schemes serve as a surety.

## Bibliography

Grosso, P. et al. (2003). *XPointer Framework*. W3C Recommendation 25 March 2003. <https://www.w3.org/TR/xptr-framework/>

TEI. (2023). P5: *Guidelines for Electronic Text Encoding and Interchange*. Version 4.6.0. Last updated on 4th April 2023, revision f18deffba. <https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html>

Sanderson, R. et al. (2017). *Web Annotation Vocabulary* W3C Recommendation 23 February 2017. <https://www.w3.org/TR/annotation-vocab/>

Cayless, H. (2013). *Rebooting TEI Pointers*. In: *Journal of the Text Encoding Initiative*. Issue 6, 2013. doi [10.4000/jtei.907](https://doi.org/10.4000/jtei.907)

## About the authors

**Ludger Hiepel** is a research assistant at the "Department of Scriptural Exegesis and Theology", University of Münster, Germany.

**Christian Lück** is a research software engineer at "Service Center for Digital Humanities", University of Münster, Germany.

**Johannes Schnocks** is a full professor at the "Department of Scriptural Exegesis and Theology" University of Münster, Germany.

## Notes

- 
1. See <https://github.com/scdh/tei-xpointer-schemes> ↵

# Text Encoding without //text: The use of //abstract as means to avoid the one-dimensionality of ego-networks in the 'Buber-Correspondences Digital' project

Denise Jurst-Görlach, Goethe University Frankfurt, Germany

Thomas Kollatz, Akademie der Wissenschaften und Literatur | Mainz

Poster

**Keywords:** correspondence, abstract, relation, network analysis

## Abstract

*Buber-Correspondences Digital*<sup>[1]</sup> is a project focused on the 41,400 preserved letters, postcards, and telegrams exchanged between Martin Buber and over 7,000 different correspondents. The goal of the project is to thoroughly research these dialogues in epistolary-form, which have received little attention thus far. A tiered editorial process is being applied using the Text Encoding Initiative (TEI). A suitable framework for the elementary metadata, such as sender, recipient, location, and date, is provided by the TEI element `<correspDesc>`, which will be made available for all correspondence units. In addition, 20% of the letters, mainly those with notable figures, are being digitally edited in full text using the TEI. 65% of the correspondences are provided with extensive structured metadata using the TEI element `<abstract>`.

Each abstract is divided into three parts: First, it contains a (human-readable) summary of essential contents organized in `<item>`s. Followed by indices listing the entities (persons, organizations, places, works, events), both those explicitly mentioned and those implicitly relevant. With regard to the planned cultural-historical analysis of Buber's dialogical relationships and networks of scholars and intellectuals represented by the correspondences, the project takes a third step: the entities are brought into relation with each other, thus breaking up the purely static presentation of a plain register. This dynamization is achieved with the help of the TEI element `<relation>`, which interrelates entities in a triple-like structure:



subject-predicate-object. By focusing on the content of letters, the 'boring' one-dimensionality that ego-networks tend to have can be overcome.

Initial evaluations of this dynamic structuralization of letter content have shown promising results, using RDF/SPARQL and graph-database Neo4j.

## About the authors

**Denise Jurst-Görlach** is a research associate at the *Buber-Rosenzweig Institute* at *Goethe University Frankfurt*. Her research interests include German-Jewish and Austrian Literature of the 19th and 20th century and Digital Scholarly Editions.

Since 2017 **Thomas Kollatz** works as a researcher in the *Academy of Sciences and Literature | Mainz*. He is a member of the *EpiDoc Development Group* and the *Digital Forum of the European Association of Jewish Studies*.

Since 2021 both are working in the long-term project (2021–2045) *Buber-Korrespondenzen Digital*.

## Notes

---

1. <https://www.adwmainz.de/projekte/buber-korrespondenzen-digital>;  
<https://agate.academy/id/PR.768> ↵

# Text-genetic studies on Carl Maria von Weber's opera fragment "Die drei Pintos"

Salome Obert, Carl Maria von Weber-Gesamtausgabe, Germany

Long Paper

**Keywords:** Weber, opera, MEI, TEI, genetic criticism

## Abstract

"I think as seldom of the Pintos as I do of music in general. I'm pretty tired of it, and I don't think I'll be doing any more work on it in the near future."<sup>[1]</sup>

Thus Carl Maria von Weber commented on the sluggish development of his opera *Die drei Pintos* in a letter of May 17th, 1824 to his friend Hinrich Lichtenstein. For months, the composer reported various people in his private and professional circle discontentedly about his lack of creative drive and the lack of time due to other engagements which kept him from working on his *comic opera in three acts* on a libretto by Karl Theodor Winkler. The date of completion was postponed several times, but finally the work grind to a halt, so that the opera remained a fragment.

Weber's surviving manuscripts of the *Pintos* can be termed as "drafts" and are today found under various signatures in the Staatsbibliothek zu Berlin – Preussischer Kulturbesitz. The most extensive of these sources, WFN 3, begins with an introduction notated as a full score. The following pages are drafts of different shape, number of voices and extension, which are partly notated as a kind of piano reduction, partly as a – rather rudimentary – orchestral piece or short score. Due to different means of writing media, corrections, notes, and other codicological characteristics, WFN 3 can be classified as working manuscript. In 1871 Friedrich Wilhelm Jaehns, author of a comprehensive catalogue of Weber's works, already remarked that this writing manner is not only specific for the compositional process of the *Pintos*, but can also be found in other manuscripts of the composer<sup>[2]</sup>. It can therefore be assumed that Weber used repeating writing strategies.

The starting point of my dissertation is the question of Weber's compositional methods examined exemplarily in the manuscripts of the opera *Die drei Pintos* followed by a digital edition of the materials – encoding the musical sources in MEI

and the libretto in TEI. Regardless of the fragmentary product of writing, Weber's writing process will be reconstructed asking for Weber's compositional approach: Can individual writing operations be identified or rather working routines? In which way are these routines subject-specific? In order to answer these questions, the WFN 3 source will be analyzed with the methods of genetic criticism developed by the research project *Beethovens Werkstatt*. In addition, Weber's diaries and his correspondence will be consulted as well: What is the relationship between those written statements about his compositional process and the writing traces found in WFN 3?

Finally, an adequate presentation of the results of the work will be the subject of the dissertation as well. In doing so, the writing traces solidified in the manuscript are to be redynamized and contextualized within the creative process. The dissertation therefore has a dual alignment: on the one hand, it wants to contribute to the elucidation of text-genetic processes and working strategies, but on the other hand, it also wants to investigate new methods for their presentation.

In my paper which is to be presented at the Joint MEC and TEI Conference 2023 I will give an overview of the compositional project *Die Drei Pintos* and its sources.

## Bibliography

Appel, Bernhard R. "Zum Textstatus von Kompositions-Skizzen und -entwürfen.", *Jahrbuch des Staatlichen Instituts für Musikforschung Preußischer Kulturbesitz*, 1999, 177–210.

Appel, Bernhard R. "Über die allmähliche Verfertigung musikalischer Gedanken beim Schreiben." 2003, *Die Musikforschung*, 56, no. 4 (2003): 347–65.

Appel, Bernhard R., and Reinmar Emanns, eds. *Musikphilologie. Grundlagen – Methoden – Praxis*. Vol. 3. *Kompodium Musik*. Laaber, 2017.

Bartlitz, Eveline, and Frank Ziegler. "'Mit den Pintos ist es ein eigen Ding ...': Das Schicksal von Carl Maria von Webers unvollendeter Oper 'Die Drei Pintos' von 1826 bis 1852 aus dem Blickwinkel seiner Familie." In *Weberiana*, 27:42–92, 2017.

Danuser, Hermann. "Urheber und Mehrer: Aspekte der Autorschaft Bei Weber/Mahlers komischer Oper 'Die Drei Pintos'." In *Weber – Jenseits Des 'Freischütz'*. *Referate des Eutiner Symposiums 1986 anlässlich des 200*.

*Geburtstages von Carl Maria von Weber*, edited by Friedhelm Krummacher and Heinrich W. Schwab, 41–58. Kieler Schriften zur Musikwissenschaft 32. Kassel, 1989.

Henze-Döhring, Sabine. *Giacomo Meyerbeer. Briefwechsel und Tagebücher*. Vol. 5. 1849–1852. Berlin, 1999.

Heusgen, Birgit. *Studien zu Gustav Mahlers Bearbeitung und Ergänzung von Carl Maria von Webers Opernfragment 'Die Drei Pintos'*. Kölner Beiträge Zur Musikforschung 133. Regensburg, 1983.

Jaehns, Friedrich Wilhelm. *Carl Maria von Weber in Seinen Werken. Chronologisch-Thematisches Verzeichniss seiner Sämmtlichen Compositionen nebst Angaben der unvollständigen, verloren gegangenen, zweifelhaften u. untergeschobenen m. Beschreibung der Autographen, Angabe der Ausgaben u. Arrangements*. Berlin, 1871.

Konrad, Ulrich. *Mozarts Schaffensweise. Studien zu Werkautograph, Skizzen und Entwürfen*. Göttingen, 1992.

Veit, Joachim, and Frank Ziegler, eds. *Carl Maria von Weber: "...wenn Ich Keine Oper unter den Fäusten habe, ist mir nicht wohl". Eine Dokumentation zum Opernschaffen*. Katalog zur Ausstellung der Staatsbibliothek Zu Berlin – Preussischer Kulturbesitz. Staatsbibliothek zu Berlin – PK. Ausstellungskataloge 45. Wiesbaden, 2001.

Weber, Carl von. "Karl Maria v. Webers unvollendet hinterlassene Komische Oper. Die drei Pintos." *Neue Zeitschrift Für Musik* 55, no. 83 (January 11, 1888): 20.

Zychowicz, James Leo. "Mahler's Instrumental 'Entr'act' for 'Die Drei Pintos' and his Emerging Symphonic Style." *Ad Parnassum* 2, no. 3 (2004): 7–30.

Diaries and correspondence of Carl Maria von Weber edited by the Carl-Maria-von-Weber-Gesamtausgabe, available at <https://weber-gesamtausgabe.de/>.

## Manuscripts

D-B, Mus.ms.autogr. Weber, C. M. v., WFN 3

D-B, Weberiana Cl. I, 24

D-B, Weberiana Cl. III, vol. 5

## About the author

**Salome Obert** studied cultural sciences and musicology in Hildesheim and Paderborn/Detmold with semesters abroad in Italy and Portugal. While working in the project *Beethovens Werkstatt* she wrote her master thesis on an analysis of scriptural problems in Beethoven's autograph to his *Flohlied* op. 75/3. Currently she is working at the Carl-Maria-von-Weber-Gesamtausgabe and writing a dissertation on Weber's opera fragment *Die drei Pintos*.

## Notes

---

1. Translated by the abstract's author, original wording: „An die Pintos denke ich so wenig jetzt, als überhaupt an Musik. habs recht satt, und werde wohl sobald keine größere Arbeit vornehmen.“ ([WeGA: A042296](#)). ↩
2. Cf. Jähns 1871: 421. ↩

# The Culture of Devotion – And what TEI Encoding tells us about it

Susanne Haaf, Berlin-Brandenburg Academy of Sciences and Humanities, Germany

Long Paper

**Keywords:** Corpus analysis, TEI corpora, German devotional literature

## Abstract

### Introduction

In text linguistics, among other things we study the structuring of texts and the question if certain text types are structured in a certain, typical way. However, even though TEI encoded texts have been around for quite some time, now, in historical German linguistics works on exploiting TEI encoding in the context of text type analysis are rare.

In my dissertation thesis (Haaf: appearing), I studied the presence of typical language patterns in certain devotional text types, namely protestant funeral sermons and devotional prose texts of the 17<sup>th</sup> century. Comparing corpora, I extracted statistically relevant textual patterns of the named text types and asked for their respective functions for the text. My work also included an exploitation of TEI encoding – alone and in combination with potential textual patterns, which I plan to report on in my talk.

The thesis combined qualitative and quantitative methods: potentially significant features were gathered from previous qualitative research and were extracted from corpora using (semi-)automatic methods. The results, however, did not just allow for insights on relevant textual features but for conclusions based on these features on the specifics of devotional 17<sup>th</sup> century culture.

### Related work

Though devotional literature was in its time highly relevant for people of all social ranks, it has only rarely been considered by linguistic, literary, or theological research (exceptions: esp. Pfefferkorn 2005; Kemper 2015). Moreover, despite vast qualitative research on the characteristics of text types in Germanic linguistics

alone (overviews e.g. Heinemann 2000; Schuster 2017; bibliography: Adamzik 1995), the approach of exploiting large corpora in this context is relatively new.

Corpus linguistics considering text types focused on contemporary texts and the differentiation of widely disparate text types (e.g. Biber 1988), or on corpus driven methods (e.g. Scharloth 2018), one exception being Bubenhofer and Spieß 2012. Digital literary studies typically base text type differentiation on stylometry or topic modelling methods rather than linguistic features (e.g. Schöch 2017; Hettinger et al. 2015; cf. Viehhauser 2017). All such studies almost never exploit TEI encoding in the corpora used.

## Method & Corpora

The current study was based on three TEI corpora (table 1), taken from the DTA (2007–2023) collections. Those texts are encoded according to the DTA Base Format (DTABf), a TEI P5 dialect which is meant to allow for homogeneous annotation and interoperable outcome of historical texts (DTABf since 2011; Haaf, Geyken, Wiegand 2014). Thus, information on text structures were available for data analysis, but also on linguistic features of tokens. The latter are gained by automatic procedures within the digitization and publication workflow of the DTA (Jurish 2012), and one output format includes `tei:w` elements with these information on token level based on the TEI `att.linguistic` class (Bański, Haaf, Mueller 2017). Feature extraction in the current study was based on this resulting format (DTABf + `att.linguistic`)<sup>[1]</sup> and was done using XSLT and Python technology.

Table 1: Corpora		
Corpus	Documents	Token (in million)
Devotional prose	35	6.9
Funeral sermons	284	3.7
Diverse text types for comparison	187	22

Table 1

The study included nineteen features (see examples in table 2) from different textual layers (word, phrase, sentence, text), estimating significance by computing and evaluating measures of descriptive and analytical statistics (on frequency, distribution, and variance).

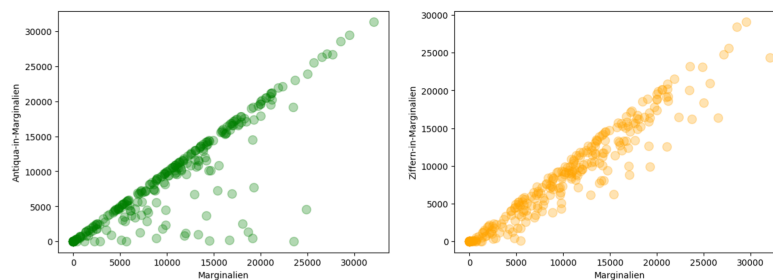
Table 2: Selected Results		
Type of Pattern	Pattern (Operationalization)	Methodological Particularity
Repetitions	Repetition of lemma bigrams & trigrams per paragraph (tei:p), heading (tei:head) and line group (tei:lg)	Factoring in paragraph length
Coordinations	Syndetical and asyndetical coordinations of similar parts of speech	Handling of false positives/negatives from automatic POS analysis
Intertextuality	1. Bibliographic references 2. Citations from Luther's New Testament translation	Absence of particular TEI encoding for most citations (as tei:cit/tei:quote) and bibliographic references (as tei:bibl)
Negation	Usage of negation markers	Usage of sentence extraction results to additionally count in negation solving with „sondern“ („but“)
Text Structuring	Presence of indices, figures, marginal notes summarizing content; heaviness of inline and document structuring in general	Semiautomatic (as opposed to automatic) approach due to inconsistencies of TEI encoding

Table 2

## Results to present

The results show, that essential information on patterns of text types can be conveyed by TEI text structuring. This concerns layout specifics as well as typical combinations of textual structures and ways of phrasing.

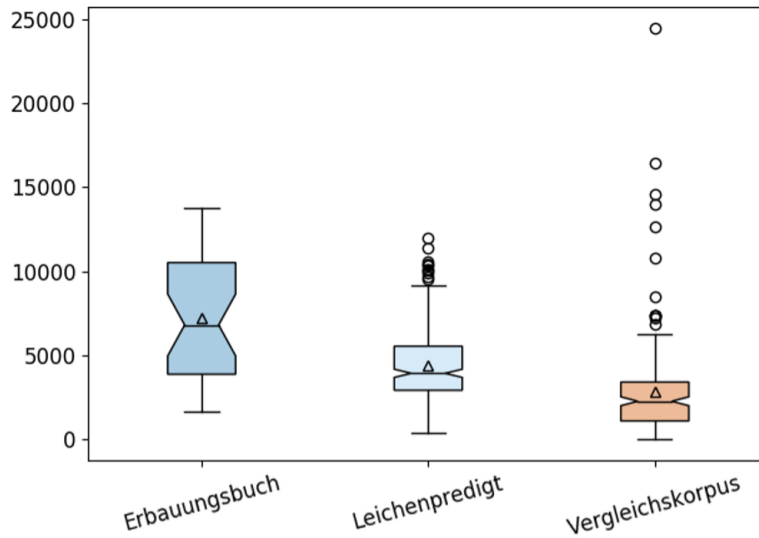
Thus, results from earlier qualitative analyses could now be specified by factoring in TEI encoding. For example, the finding that citations are essential to devotional literature could now be supplemented with information about how and where bibliographic citations are usually realized in a text (Img. 1).



Img. 1: Margins in funeral sermons carrying (typographic characteristics of) bibliographic citations, freq. per 1 million token

Furthermore, characteristic places of lexical repetition (Img. 2) could be specified along with the relevance of location for its emotionalizing effect (repetitions in lists vs. paragraphs).





Img. 2: Repetitions of trigrams in paragraphs in the devotional prose (left box), funeral sermons (middle box) and the reference corpus (right box), freq. per 1 million token

These are only examples of a range of results obtained by considering TEI encoding, which I would like to present. However, the talk shall also address limitations of the approach (i.e. limited annotation depth or lacks of interoperability as briefly listed in table 2), ways to stretch these limits, and requirements for markup depth in linguistic research data.

Finally, a model gained from the study's results will be presented that shows intended effects of devotional literature and its significant textual and structural patterns, and thus allows for insights on German devotional culture of the 17<sup>th</sup> century.

## Bibliography

Adamzik, Kirsten (1995): *Textsorten - Texttypologie. Eine kommentierte Bibliographie* (Studium Sprachwissenschaft 12). Münster.

Bański, Piotr, Susanne Haaf & Martin Mueller (2018): 'Lightweight Grammatical Annotation in the TEI. New Perspectives'. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, 7.-12.05.2018, Miyazaki (Jp), 1795-1802. Online: <https://aclanthology.org/L18-1283.pdf> (accessed 2023-06-20).

Biber, Douglas (1988): *Variation Across Speech and Writing*. 1st ed. (digital version of 2012). Cambridge et al. DOI: [10.1017/CBO9780511621024](https://doi.org/10.1017/CBO9780511621024).

Bubenhofer, Noah, and Constanze Spieß (2012): 'Zur grammatischen Oberflächenstruktur von Kommentaren: Eine korpuslinguistische Analyse typischer Sprachgebrauchsmuster im kontrastiven Vergleich'. In: *Presstextsorten jenseits der "News". Medienlinguistische Perspektiven auf journalistische Kreativität*, ed. by Christian Grösslinger, Gudrun Held, and Hartmut Stöckl (Sprache im Kontext 38), Frankfurt am Main, pp. 85–105.

DTA (since 2007): *Deutsches Textarchiv. Grundlage für ein Referenzkorpus der neuhochdeutschen Sprache*. Prepared by Matthias Boenig, Alexander Geyken, Susanne Haaf, Bryan Jurish, Christian Thomas, and Frank Wiegand. Ed. by Berlin Brandenburg Academy of Sciences and Humanities. Berlin. Online: <https://deutschestextarchiv.de/> (accessed 2023-07-31).

DTABf (since 2011): *Deutsches Textarchiv – Basisformat*, ed. by Deutsches Textarchiv (DTA) and DTABf Steering Committee (Susanne Haaf, Matthias Boenig, Alexander Geyken, Christian Thomas, Frank Wiegand, Daniel Burkhardt, Stefan Dumont & Martina Gödel). Berlin. Online: <https://deutschestextarchiv.de/doku/basisformat> (accessed 2023-06-22).

Haaf, Susanne (in preparation): *Die Musterhaftigkeit erbaulicher Textsorten des Deutschen. Ein integrativer Ansatz zu ihrer Untersuchung an der Schnittstelle von quantitativer und qualitativer Linguistik*. Diss., Univ. Paderborn, 2022 (to appear 2023).

Haaf, Susanne, Alexander Geyken, and Frank Wiegand (2014): 'The DTA "Base Format". A TEI Subset for the Compilation of a Large Reference Corpus of Printed Text from Multiple Sources'. In: *Journal of the Text Encoding Initiative (JTEI)* 8. DOI: [10.4000/jtei.1114](https://doi.org/10.4000/jtei.1114).

Heinemann, Wolfgang (2000): 'Aspekte der Textsortendifferenzierung'. In: *Text- Und Gesprächslinguistik. Ein internationales Handbuch Zeitgenössischer Forschung*, ed. by Klaus Brinker, Gerd Antos, Wolfgang Heinemann, and Sven F. Sager (Handbücher zur Sprach- und Kommunikationswissenschaft 16.1), Berlin/New York, pp. 523–546.

Hettinger, Lena, Martin Becker, Isabella Reger, Fotis Jannidis, and Andreas Hotho (2015): 'Genre Classification on German Novels'. In: *Proceedings of the 26th*

*International Workshop on Database and Expert Systems Applications (DEXA)*, IEEE, pp. 249–253.

Jurish, Bryan (2012): *Finite-State Canonicalization Techniques for Historical German*. Diss., Univ. Potsdam, 2011. Potsdam. Online: <http://opus.kobv.de/ubp/volltexte/2012/5578> (accessed 2022-06-14).

Kemper, Karl-Friedrich (2015): *Religiöse Sprache zwischen Barock und Aufklärung. Katholische und protestantische Erbauungsliteratur des 18. Jahrhunderts in ihrem theologischen und frömmigkeitsgeschichtlichen Kontext*. Diss., Philosophisch-Theologische Hochschule SVD St. Augustin (Religionsgeschichte der frühen Neuzeit 22). Nordhausen.

Pfefferkorn, Oliver (2005): *Übung der Gottseligkeit. Die Textsorten Predigt, Andacht und Gebet im deutschen Protestantismus des späten 16. und 17. Jahrhunderts*. Habil., Univ. Halle-Wittenberg 2003 (Deutsche Sprachgeschichte 1). Frankfurt am Main.

Scharloth, Joachim (2018): 'Korpuslinguistik für sozial- und kulturanalytische Fragestellungen'. In: *Korpuslinguistik*, ed. by Marc Kupietz and Thomas Schmidt (Germanistische Sprachwissenschaft um 2020 5), Berlin/Boston, pp. 61–80.

Schöch, Christof (2017): 'Topic Modelling Genre. An Exploration of French Classical and Enlightenment Drama'. In: *Digital Humanities Quarterly* 11 (2). Online: <http://www.digitalhumanities.org/dhq/vol/11/2/000291/000291.html> (accessed 2022-02-11).

Schuster, Britt-Marie (2017): 'Elemente einer Theorie des Textsortenwandels. Eine Bestandsaufnahme und ein Vorschlag'. In: *Textsortenwandel vom 9. bis zum 19. Jahrhundert. Akten zur internationalen Fachtagung an der Universität Paderborn, 9.–13.6.2015*, ed. by Britt-Marie Schuster and Susan Holtfreter (Berliner sprachwissenschaftliche Studien 32), Berlin, pp. 25–43.

Viehhauser, Gabriel (2017): 'Digitale Gattungsgeschichten. Minnesang zwischen generischer Konstanz und Wende'. In: *Zeitschrift für digitale Geisteswissenschaften* (2). DOI: [10.17175/2017\\_003](https://doi.org/10.17175/2017_003).

## About the Author

**Susanne Haaf** holds a degree in German philology and Computational Linguistics (M. A.) from the University of Heidelberg. Currently, she works as a research

associate at Berlin-Brandenburg Academy of Sciences and Humanities, where she has been engaged in the projects [DTA](#), [CLARIN-D](#), [t.evo](#) and (till present) [ZDL](#), all of which involved the preparation and maintenance of TEI corpora. She finished and defended her PhD thesis at the University of Paderborn in 2022, which contains work on the computational analysis of patterns which differentiate historical devotional text types.

## Notes

---

1. I am grateful to my former colleague Bryan Jurish who added this format to the selection of DTA download formats. [↩](#)

# The Past, Present, and Future of the Plaine and Easie Code

Andrew Hankinson, RISM Digital Center, Switzerland

Jennifer Ward, RISM Editorial Center, Frankfurt

Long Paper

**Keywords:** PAE, notation, encoding, bibliography

## Abstract

The *Plaine and Easie Code System* (PAE) for music notation had its beginnings in the 1960s as a means of communicating music notation using alphanumeric characters. Its primary use has been in the capture of “incipits”, or the first few opening measures of a musical piece for the purposes of positive identification of a work based on its musical content. Since its introduction it has undergone several revisions and modifications, and adoption by the international library community. Today it sees daily use in the Répertoire International des Sources Musicales (RISM) project as the basis for capture of thematic incipits, which currently number over 1.4 million individual entries.

In 1964, Barry S. Brook and Murray Gould published *Notating Music with Ordinary Typewriter Characters (A Plaine and Easie Code System for Musicke)*. This was the first paper outlining this coding scheme. In the years that followed, Brook presented the code at international meetings and promoted PAE through publications. Throughout, he sought to reach people in both musicology and librarianship while appealing to the small but growing community interested in computer applications for the humanities. He continuously appealed to RISM—then, as now, the world's largest cataloguing project for written music—to include music incipits in their inventories. In the early 1970s Brook succeeded in convincing RISM to adopt PAE for music incipits. From here on, RISM's cataloguing efforts steered all further publications and dissemination of the code.

In 2022, the RISM Digital and Editorial Centers launched RISM Online (<https://rism.online>), a new way to search the RISM data and featuring a new incipit search system. This incipit search provides notation-centric search facilities, both with a keyboard input system and through direct PAE input as a query language. As

part of making this data more publicly visible, it was necessary to undertake two projects to help improve data quality for the RISM incipits and the digital tools that enable their use in search and display.

The first project, in collaboration with the Centre for Digital Music Documentation at the Academy of Sciences and Literature in Mainz, sought to improve consistency and overall data quality in the RISM incipit data. The resulting correction work highlighted problems in the PAE specification itself, so a second project focused on making improvements in this area. This has taken the form of editorial improvements aimed at making the PAE code easier for software developers and other digital tool makers to use this notation format, correctly, in their own system, through the use of normative language, clear examples, and some backwards-incompatible changes to the encoding schema. This will culminate in a new version of the PAE specification.

Our talk will cover a history of this innovative, but often overlooked, music notation encoding scheme, our current efforts to improve it. We will present the results of our projects for improving RISM incipit data, and discuss new tools, applications, and directions for this venerable technology.

# The Ricercar Data Lab score viewer

Suzy Piat, Ricercar Lab, Centre of Renaissance Studies, University of Tours, France

Poster

**Keywords:** database, PostgreSQL, Django, Verovio, MEI

## Abstract

The research programme in musicology Ricercar Lab, at the Centre of Renaissance Studies (University of Tours / National Centre for Scientific Research, France), has gathered since its creation a lot of data related to different research projects, stored in several databases and displayed through individual websites. In 2020, a single relational database especially thought for musical data was built, with a dedicated website, using the database management system PostgreSQL and the Python-based Web framework Django. The database focuses on three main entities – musical works, sources and people – and the website allows users to browse the data and to perform search queries. The data is enriched with links to external resources, such as digitisations of the sources, authority files like the VIAF and references to the RISM catalog. The available scores can be downloaded in different file formats, MEI and PDF in particular.

A dedicated interface allows users to browse scores within a particular source or work. Each score can be seen and played according to several options, regarding the score display (orientation, zoom, etc.) or the audio (instrument, tempo, etc.). This interface uses the music notation engraving library Verovio, developed by the RISM Digital Center, to display the score as SVG. A side section allows to view either the MEI code or the critical apparatus encoded in the MEI file. Annotations and variations are displayed according to the measure they belong to, with score excerpts to view variations. Links are made between the score, the MEI code and the annotations to highlight a particular item. Logged-in users are allowed to suggest annotations related to none, one or several items in the score, using a form with fields for content and type. Annotations are then reviewed by the Ricercar Lab team and inserted into the MEI file if validated.

## About the author

**Suzy Piat** studied art history, digital humanities and IT and she specialised in web development, in relation to databases. She works with the research programme in musicology *Ricercar Lab*, at the Centre of Renaissance Studies (University of Tours, France), and she is in charge of the development of the website dedicated to the *Ricercar Lab* database.



# The WizKit: A New Environment for Editing and Presenting MEI Metadata

Clemens Gubsch, Austrian Academy of Sciences (ÖAW), Austria

Paul Gulewycz, Austrian Academy of Sciences (ÖAW), Austria

Peter Provaznik, Austrian Academy of Sciences (ÖAW), Austria

Short Paper

**Keywords:** Metadata, Tool Development, static HTML

## Abstract

The input, presentation and searchability of music-related metadata is a major challenge for many musicological projects, especially for people still mostly unfamiliar with DH workflows. Editing with XML, particularly in the multi-layered area of the MEI header, with the highest possible uniformity for output on a website, can sometimes prove to be as complex as the encoding of the content itself. Adding searchability and allowing the combination of different search parameters complicates matters even further.

Although MerMEId is a tool for the three areas of the data cycle mentioned above, the application does not always accommodate project-specific requirements. At the same time MerMEId's associated installation, service effort and overall learning curve for using the tool can seem disproportionate for entering rudimentary work, source and performance data.

This is where WizKit comes in: WizKit implements a project structure including a framework for the Oxygen XML Editor that can be used for data input in author mode (Oxygen Form Controls), data presentation in the web browser (generated static HTML) as well as for full-text search and faceting (pre-indexed search with Typesense). Work, person/institution and literature data are entered with predefined forms, which are configured using CSS and can be easily customised. The static HTML output is generated by means of XSLT and enables the editor to straightforwardly review the entered data. The WizKit environment provides the users with ready-made functionalities with numerous possibilities for customisation.

Experimentally, the relevant MEI encodings are created from the manuscript metadata in combination with image files; these encodings then serve as the basis

for generating IIIF manifests, including tables of contents (derived from <content>. Usability of the manuscripts is improved via a local IIIF viewer.

# Thinking about collation

**Elisa Eileen Beshero-Bondar**, Penn State Erie, The Behrend College, United States of America

**Raffaele Vigiante**, University of Maryland

**Hugh Cayless**, Duke University, United States of America

**Torsten Roeder**, Universität Würzburg, Germany

**Workshop**

**Keywords:** collation, critical apparatus, variance, automation, TEI and MEI

## Abstract

Collation projects, or projects that attempt to compare and model versions, pose distinct challenges for the scholarly conceptualization of a text. The critical apparatus encoding tagsets available in the TEI and the MEI offer structures to help model variation in multiple ways, but the explanation of these could be improved and those engaged in collation projects may wish to improve the examples and possibilities for modeling available in our community guidelines. The leaders of this workshop have each written about, modeled, processed, and otherwise experimented with critical apparatus and collation methods. Each brings a distinct perspective on problems and possibilities for computational modeling of alignment and variation, as well as strategies to process collations for analysis and scholarly publication.

While the workshop leaders are primarily knowledgeable of the TEI, the workshop is directed toward members of the TEI and MEC communities and, depending on the participants, may explore differences in the collation of music notation vs. text-based documents. We are best prepared to discuss textual collation, but we invite music notation experts to participate since both communities share related data models for the critical apparatus and as we work with multimodal documents combining music notation and text. The workshop leaders will share a diverse array of experiences with collation problems, with emphasis on finding solutions in conceptual modeling of texts and variation. Participants will be expected to bring their own laptops, and we will need a projector in the room.

We hope that this workshop will help to formulate new ideas about how to model variations. While tools and methods for collation are by necessity distinct for text

and music notation, the results share notable commonalities, as demonstrated by the similarity of critical apparatus tags in TEI and MEI. Moreover, the workshop will aim to identify which areas can be improved in our Guidelines (both TEI and MEI) on this subject. For example, in the TEI Guidelines Chapter 12, the discussion of location-referenced double-endpoint-attachment, and parallel segmentation methods may unnecessarily complicate the question of whether they can be encoded in-line or externally, and whether variations can be represented as overlapping or not. Further, the TEI's examples prioritize the highlighting of differences among similar witnesses. Perhaps we should consider an alternative, to model similarities in closely related but heavily variant texts.

Of course, the TEI Guidelines merely suggest without explicating how computational processes can work with the critical apparatus to reconstruct witnesses from the encoding and why that may be a desirable goal. Our workshop will try to explore, identify, and articulate the gaps between the data model and the presumptions of the tooling we have applied.

While the word "collation" does not appear in the current version of Chapter 12 of the TEI Guidelines, "collation" is now more frequently paired with "machine-assisted" or applied in the context of automation. Given the importance of computational tooling to the workflow of collation, particularly when it comes to text or music processing, we seek to explore the special challenges involved in trying to generate good output: what fine-tuning methods can we apply, what strategies are brittle or problematic? The MusicDiff tool developed for *Beethoven's Werkstatt* aligns and visualizes music scores marked in a simple form of MEI, and as discussed at the 2020 MEI conference, its developers anticipate that it could be generalized to a wider range of encoding.<sup>[1]</sup> For text collation, software such as collateX is often used to help prepare the foundation of a TEI critical apparatus, but the application of such software and how we prepare its inputs and outputs requires careful consideration. The TEI Guidelines do not recommend a serialization for automated collation, but perhaps they ought to. The *Gothenburg Model*, first formulated in 2009 for text collation, helps to establish a method for thinking about the various procedures involved: tokenization, normalization, alignment, analysis/feedback, and visualization.<sup>[2]</sup> Significantly, these do not proceed in a linear way. Visualization exposes problems and we have to revisit our paradigms for tokenization and normalization based on analysis of the results, and especially challenging collation projects may seem to require endless adjustments. Workshop participants new to collation will gain familiarity with the *Gothenburg model*, and

consider how effectively its notion of alignment accords with our notions of segmentation.

## Outline of the workshop

1. Collation in theory: an introduction
2. Survey of established methods:
  - a. Machine-assisted collation and the Gothenburg model for textual data
  - b. Critical apparatus modeling
  - c. TEI vs. MEI approaches
3. What can we do with collation data that is not well represented in our current community Guidelines in TEI and MEI?
4. Problems we experience with collation: Workshop leaders and participants share representative examples of conceptual and computational challenges for collation for identifying alignments and variations in textual data.
5. Experimental/Open Discussion: Can AI-training support collation?
6. Sharing and workshopping collation use-cases and plans.

## About the authors

**Elisa Beshero-Bondar** is Professor of Digital Humanities and Program Chair of Digital Media, Arts, and Technology at Penn State Erie, The Behrend College. Her work on the Frankenstein Variorum project has led her into some interesting challenges with machine-assisted collation and the TEI critical apparatus. She is chair of the TEI Technical Council, on which she has served as an elected member since 2016.

**Raffaele (Raff) Viglianti** is a Senior Research Software Developer at the Maryland Institute for Technology in the Humanities, University of Maryland. His research is grounded in digital humanities and textual scholarship, where “text” includes musical notation. He researches new and efficient practices to model and publish textual sources as innovative and sustainable digital scholarly resources. He is currently an elected member of the Text Encoding Initiative technical council and the Technical Editor of the Scholarly Editing journal.

**Hugh Cayless** is a Senior Digital Humanities Developer at Duke University Libraries. His focus in the digital critical edition space has been on improving the TEI Guidelines' treatment of textual variation issues and on developing interactive visualizations of critical apparatus. His work concentrates mainly on ancient texts, including papyri and inscriptions. He has re-edited chapter 12 (Critical Apparatus) of the Guidelines extensively, but still finds problems with it every time he looks at it.

**Torsten Roeder** is a Senior Digital Humanities project manager at the Centre for Philology and Digitality at University of Würzburg. His main area are digital scholarly editions and he worked for various projects that focus on textual genetics, variance and comparison. His team is developing a generic interface framework for digital resources, while his own research project deals with early born-digital heritage and semantics of digitally represented text.

## Notes

---

1. Kristin Herold, Johannes Kepper, Ran Mo, and Agnes Seipelt, "MusicDiff – A Diff Tool for MEI," *Music Encoding Conference Proceedings*, eds. De Luca, E. & Flanders, J. 2020 59-66. Humanities Commons. [↵](#)
2. For a detailed explanation of the Gothenburg Model, see Interedition Development Group, *The Gothenburg Model*, 2010-2019: <https://collatex.net/doc/>. On the summit and workshop of collation software developers in 2009 that formulated the Gothenburg model, see Ronald Haentjens Dekker, Dirk van Hulle, Gregor Middell, Vincent Neyt, and Joris van Zundert, "Computer-supported collation of modern manuscripts: CollateX and the Beckett Digital Manuscript Project," *Digital Scholarship in the Humanities* 30:3 (December 2014) pp. 3-4. DOI: <https://doi.org/10.1093/lc/fqu007>. [↵](#)

# “This type of data in the header” – Encoding, (re)use and standardization of metadata in the context of TEI and MEI

**Dennis Ried**, University of Paderborn

**Sophia Dörner**, Sächsische Landesbibliothek Staats- und Universitätsbibliothek  
Dresden, Germany

**Clemens Gubsch**, Austrian Academy of Sciences (ÖAW), Austria

**Mark Saccomano**, Paderborn University, Center for Music, Edition, Media  
(ZenMEM)

**Panel**

**Keywords:** metadata, standardization,  
interoperability

## Abstracts

### Introduction

In the field of digital humanities working with TEI or MEI or both a section specifically for metadata is available: the header. Metadata is a wide field but always present, however, this does not mean that metadata is always encoded. In recent years, the use of digital technologies has become more and more important in digital humanities research, e.g., by creating applications to enhance research in traditional humanities. One of the challenges in the field of digital humanities is the management of metadata. This type of data is important for organizing and interpreting digital resources. The challenges arise from the requirements and the opportunities of metadata in this context. Metadata should help to identify the resources, should be research data on a high level, could be processable (well-structured metadata) which promotes findability, reusability (e.g., analysis), or simply the identification of the resource.

The idea for this panel was born in a meeting of the MEI Cataloging and Metadata Interest Group. In monthly meetings, the group works on documentation and improvements concerning metadata in MEI. The rich structure of the MEI-Header, especially the section for source descriptions, is heavily based upon TEI but also provides support for additional information that TEI does not, for example,

elements that allow representation according to the FRBR conceptual reference model. But it is not all about the differences between TEI and MEI. Regardless of any difference between the two standards, the problems we encounter concerning metadata in the digital humanities are often similar but are rarely addressed in an interdisciplinary manner.

- Who is the expected target audience of the encoded metadata? (e.g., libraries/digital archives, data analysis, search engines)
- Where and how could we share the metadata? (e.g., correspSearch, metadata exchange format, restAPI?)
- Perspectives and models for sharing and integrating the different standards? (e.g., transformation MEI-metadata to TEI-metadata and vice versa)
- What are the challenges in the creation and management of metadata?
- Should we aim for more standardization, e.g., by providing best practice recommendations?
- Is the creation of (more) structured metadata a desirable goal?

Metadata plays a critical role in the field of digital humanities as it facilitates organization, distribution and interpretation of digital resources. The management and standardization of metadata is essential for efficient and effective research as well as interoperability. This panel will reflect on the differences and similarities of handling metadata in the scope of TEI and MEI. Bringing together researchers from both communities is a much-needed corrective to this issue.

## **Structure of the panel**

The panel will be an open round table introduced by short talks. The spotlight talks should initialize the discussion on certain topics. After giving additional input by the (4–5) panelists the discussion should be opened directly, so the audience can participate early. This is very important because the group of panelists is not as international as it could be and the panel will profit from perspectives and insights coming from outside.

The expected outcome of this panel is to bring together members of both the TEI and MEI communities who face the same or similar challenges in their respective domains. It is meant as an opportunity to get in contact, discuss or organize further work on this topic in an interdisciplinary context.

The following people have stated their active participation (context in brackets):

- Margrethe Støkken Bue (MEI, Library)
- Peter Stadler (TEI, Research Software Engineering)



- Magdalena Tuska/Wolfgang Meier (TEI, Research Software Engineering)
- Barbara Wiermann (Library)
- Mark Saccomano (MEI)
- Nikolaos Beer (MEI/TEI, Research Software Engineering)

# Tools, Formats, Mappings: Encoding Franz Liszt's Œuvre between MARC and MEI

Severin Kolb, SLUB Dresden

Matthias Richter, SLUB Dresden

Long Paper

**Keywords:** Digital Musicology, Catalogue of Works, MerMEId, RISM

## Abstract

Franz Liszt's oeuvre – despite its undeniable central importance for the music history of the 19th century – is still insufficiently researched today. His manifold activities and works polarized not only during his lifetime, but also in its posthumous reception and in research. A cosmopolitan musician and composer, Liszt was located in a vacuum between the increasingly nationally oriented music historiographies. This is still reflected in today's research, which lacks important foundations: Historical-critical editions of works are only gradually emerging. Editions of his extensive correspondence are outdated and unreliable. And last but not least, the lack of a reliable catalogue raisonné of works and sources hinder research.

With the development of an online portal, the DFG project *Digital Liszt Source and Work Catalogue* (University of Heidelberg, Goethe and Schiller Archive Weimar, SLUB Dresden) is laying a new foundation for source-based Liszt research. The portal will host the catalogue raisonné and be able to become an umbrella for future projects focused on Liszt related sources as well. All projects will communicate via central data services, providing the research community with a plethora of well searchable and navigable data on music, writings, performances, research literature, and the like. The portal aims to serve an international research community by providing an internationalised user interface and localisations of metadata. All data will be available via standardised APIs. Central services include registers, bibliographic data sets, a search engine and databases for relational and XML data. The portal will encourage the usage of large research infrastructure projects such as RISM, authority files, musicconn.performance or Zotero. It will evaluate different kinds of data integration via a search engine only vs. approaches of translating data sets from third party repositories to formats which may be stored locally like MEI or TEI.

The project focusses on Liszt's musical oeuvre, which constitutes the portal's starting point. It poses considerable challenges, as many of his works evolve during a long process characterized by constant transformation resulting in different versions of a work or closely related work groups. Additionally, improvisational practices of Liszt's time, profoundly shaping his oeuvre, and the collaboration with other composers, musicians, scribes, and publishers lead to interesting but quite complex relations between works and their expressions.

As a result of this creative process, a plethora of sources, often interconnected, are found in libraries and archives all over the world. For generations, scholars have tried to understand the nature of this oeuvre and to find a system to describe it aptly. From the same starting point, researchers have come up with a surprisingly broad spectrum of attempts to delineate which collections of sources constitute a specific work. No attempt to compile a complete and detailed catalogue raisonné has yet been successfully completed. With the digital turn, new foundations for scholarly work are now available. The digital medium enforces data organisation according to data models. Thus, pre-existing quality standards may be used on the one hand and pre-existing tools on the other. FRBR-based open-source tools (RISM's Muscat or MerMEId) for working on those data structures can be adapted to project-specific needs.

Given the nature of Liszt's oeuvre, the project aims to apply a „source first“ approach. It inductively infers the more abstract work data from source data. MerMEId's rather monolithic work-based data model makes this approach quite difficult. As an alternative the project seeks to store data from the FRBR layers „work“ and „expression“ separately from the levels „manifestation“ and „item“ and establish id based connections between datasets as done in relational databases. To gather lower level information and structured source data, primarily RISM's infrastructure is adapted, either queried from the RISM API and displayed directly in the Liszt Portal or translated to MEI and stored in an adapted form of the MerMEId. For information on performances and literature, which are only secondarily catalogued in RISM and with little differentiation, Zotero and MusicConn.performance are used as primary hubs for the creation of separate datasets. This approach allows for a flexibly arranged presentation of the records, which are thus ideally reusable and more in keeping with the characteristics of Liszt's oeuvre than the lists of conventional printed catalogs of works or the MerMEId front-end.

## About the authors

**Severin Kolb** and **Matthias Richter** are members of the DFG-project team *Digitales Liszt Quellen- und Werkverzeichnis* (Universität Heidelberg, Goethe- und Schiller-Archiv Weimar, SLUB Dresden)

# Toward a TEI/RDF Encoding for Semantic Annotations: Concept and Implementation as LOD Editor

Jun Ogawa, ROIS-DS Center for Open Data in the Humanities, Japan

Kiyonori Nagasaki, International Institute for Digital Humanities

Satoru Nakamura, The University of Tokyo

Ikki Ohmukai, The University of Tokyo

Asanobu Kitamoto, ROIS-DS Center for Open Data in the Humanities, Japan

Long Paper

**Keywords:** TEI, RDF, Linked Data, editor, stand-off

## Abstract

Over the years, the interoperability between TEI and external data models has been a prominent subject of discussion. Recently, the LINCS team explored the conversion from TEI to CIDOC-CRM<sup>[1]</sup>, while John Bradley introduced the TEI-based description of the Factoid Prosopography, earlier presented as RDB or RDF data<sup>[2]</sup>. In the same line, D. L. Schwartz delved into the description of Factoids in TEI<sup>[3]</sup>. In addition, ongoing theoretical discussions regarding the application of RDFa to the TEI schema can be found on the TEI GitHub issues<sup>[4]</sup>. All these studies contemplate how graph-structured data can be effectively conveyed in TEI/XML, advocating the use of TEI namespace in the LOD context. One of the main reasons behind adopting TEI/XML for LOD creation lies in its ability to represent detailed textual information more effectively than pure RDF representation<sup>[2]</sup>.

Ø. Eide's research extensively explores both theoretical and practical aspects of TEI and data models<sup>[5]</sup>. Eide perceives TEI not only as a markup method but also as a model capable of compatibility and interoperability with diverse data structures. Building upon this discussion, we propose the concept of TEI/RDF encoding. In this approach, while the basic textual structure is marked with TEI/XML, encompassing characters and editorial marks, semantics beyond the character level, such as words, entities, events, and relationships, are described in RDF using a data model compatible with the TEI namespace. These RDF elements are then linked to the

TEI/XML in quasi-standoff markup through references to @xml:id assigned to each <c> element. This linking allows for a flexible representation of the knowledge network in RDF while preserving the more effectively encoded textual features, like editorial marks, within TEI/XML (Fig. 1).



Fig. 1: Basic concept of our model

Using historical documents as a case study, we developed an ontology named HIMIKO (Historical Micro Knowledge and Ontology). The basic concept of this ontology is to capture the small pieces of historical information (Micro Knowledge in our terms), which would be about historical events, situations, or relationships, as a LOD resource encompassing the specific textual contexts and descriptions. In this sense, while rooted in the basic concept of Factoid Prosopography, this ontology expands significantly to enhance its referentiality to primary source descriptions [6]. Although HIMIKO is intended for RDF description, and therefore not XML-based, numerous classes and properties can be aligned with TEI schema concepts, as depicted in Fig. 2. Consequently, even though most elements, except <c>, aren't directly encoded in XML format, it remains largely compatible with TEI/XML encoding.

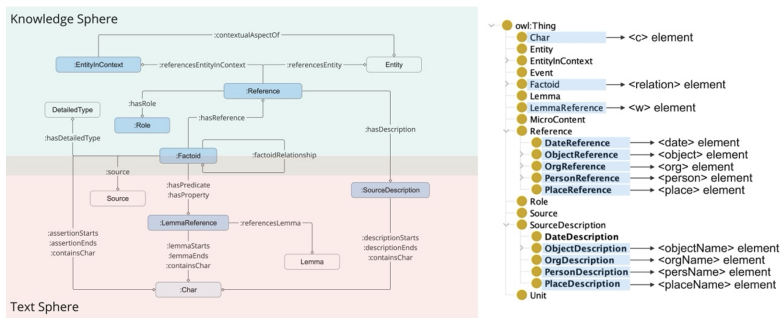


Fig. 2: The ontology and the conceptual mapping between HIMIKO and the TEI schema

The rationale behind restricting XML usage solely to textual descriptions is based on the belief that handling complex and multi-layered information is more feasible and appropriate in RDF. An illustration of this is shown in Fig. 3. Here, the issue of

overlap makes it challenging to integrate multi-layered encoding of editorial markup and entity descriptions, unless a distinct representation for entities is employed.

```
<lb n="1"/>
<expan><abbr>T</abbr><ex>ito</ex></expan> Belgis<supplied reason="lost">io<certainty locus="name"
match="preceding-sibling::text()" cert="low"/> <gap reason="lost" extent="unknown" unit="character"/></supplied> <lb n="2"/>
<supplied reason="lost">ad aram omnibus</supplied></sup><supplied reason="lost">ad aram omnibus</supplied></sup><supplied reason="lost">ad aram omnibus</supplied></sup>
<supplied reason="lost">apud suos functo</supplied> <lb n="4"/><sup>statuam</sup><supplied reason="lost"
cert="low">questrem</supplied> <lb n="5"/><sup>cum</sup><supplied reason="lost">bscriptione</supplied> <lb n="6"/><object>tres
<sup>provis</sup><supplied reason="lost">in</sup><sup>iciae</sup></object> Galliae</supplied> <lb n="7"/><sup>primo</sup><sup>u</sup><supplied reason="lost"
cert="low">quam</sup> ex? civit(ate)? sua</supplied> <lb n="8"/><sup>ponend</sup><supplied reason="lost">am</sup> curaverunt</
supplied>
```

Fig. 3: Examples of overlapping problem (original TEI/XML data from: <https://edh.ub.uni-heidelberg.de/edh/inschrift/HD000650>)

Thus, our approach considers TEI as a data model rather than a specific markup method, enabling an efficient encoding of text-related knowledge via RDF. This approach also contributes to reduced encoding costs by enabling separate processing of semantic information, potentially in a different format. It is for this purpose that we developed an editor<sup>[6]</sup>. Within this editor, the ‘Linking Field’ allows the creation of RDF data, which is basically independent of the XML syntax, based on minimally encoded TEI/XML data displayed in the ‘Text Field’ as an interactive text interface. Thus, users would not be required to manually edit either XML text markup or any RDF format, such as RDF/XML or Turtle. Our system shares similarities with CATMA<sup>[7]</sup> in its standoff implementation, and with Recogito or LEAF-Writer<sup>[8]</sup> in its data-linking system, while significantly supplementing them for more complex RDF knowledge representation.

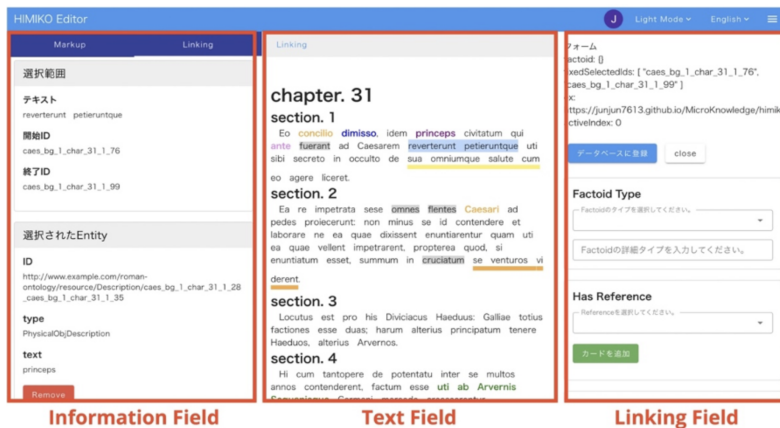


Fig. 4: Interface view of our editor

Our study aligns with previous studies that view TEI as a data model and explore its integration with Linked Data format. Building upon these works, we establish a more detailed connection between TEI encoding and knowledge description, leading to a

comprehensive implementation of Linked Data, or more specifically RDF, for representing text-related knowledge. Consequently, our study serves as a practical example of successful TEI/RDF implementation, showcasing the potential for more complex and context-dependent representation of textual information within the broader context of knowledge networks.

## Notes

---

1. C. Crompton, H. Zafar and A. Defours (2022), 'Long Paper: LINCS' Linked Workflow: Creating CIDOC-CRM from TEI,' *TEI2022 Conference Book*, pp. 114-115. [↵](#)
2. J. Bradley and D. Jakacki (2021), 'Combining the Factoid Model with TEI: examples and conceptual challenges,' poster at TEI Member's Meeting 2021, <https://hcommons.org/deposits/item/hc:42095>. [↵](#) [↵](#)
3. D. L. Schwartz, N. P. Nathan and K. Torabi (2022), 'Modeling a Born-Digital Factoid Prosopography using the TEI and Linked Data,' *Journal of the Text Encoding Initiative* [Online], Rolling Issue, <https://journals.openedition.org/jtei/3979>. [↵](#)
4. <https://github.com/TEIC/TEI/issues/1860>. [↵](#)
5. Ø. Eide (2015), 'Ontologies, Data Modeling, and TEI,' *Journal of Text Encoding Initiative* [Online], Issue 8, <https://journals.openedition.org/jtei/1191>. [↵](#)
6. J. Ogawa, I. Ohmukai, S. Nakamura and A. Kitamoto (2023) 'Collecting Pieces of Historical Knowledge from Documents: Introduction of HIMIKO (Historical Micro Knowledge and Ontology),' short paper, DH 2023, Graz/Austria, July 10-14 2023, [https://www.conftool.pro/dh2023/index.php?page=browseSessions&form\\_session=93#paperID810](https://www.conftool.pro/dh2023/index.php?page=browseSessions&form_session=93#paperID810). [↵](#) [↵](#)
7. <https://catma.de/>. [↵](#)
8. S. Brown, D. Jakacki, J. Cummings, M. Ilovan, L. Frizzera, R. Milio and C. Black (2022), 'Workshop 6: Engaging TEI Editors Through LEAF-Writer,' *TEI 2022 Conference Book*, pp. 28-30. [↵](#)



# Towards correspSearch v3.0

**Stefan Dumont**, Berlin-Brandenburg Academy of Sciences and Humanities,  
Germany

**Sascha Grabsch**, Berlin-Brandenburg Academy of Sciences and Humanities,  
Germany

**Jonas Müller-Laackman**, State- and University Library Hamburg

**Ruth Sander**, Berlin-Brandenburg Academy of Sciences and Humanities, Germany

**Steven Sobkowski**, Berlin-Brandenburg Academy of Sciences and Humanities,  
Germany

## Long Paper

**Keywords:** correspondence, correspDesc,  
CMIF, correspSearch, web service

The web service correspSearch aggregates correspondence metadata from digital and printed editions (or scholarly catalogues of correspondence) and offers them for central search and retrieval (Dumont 2016). The web service is based mainly on the Correspondence Metadata Interchange Format (CMIF), which is being developed within the TEI Correspondence SIG (Stadler 2014; TEI Correspondence SIG 2018). The CMIF is built on the TEI element correspDesc (Stadler, Illetschko, Seifert 2016), which is used in a highly reduced and restrictive way. URLs from authority files are extensively used for the identification of persons and places (cf. Stadler 2012).

The web service correspSearch was made available as a prototype in 2014. Since 2017, a new software architecture, funded by the Deutsche Forschungsgemeinschaft (DFG), has been implemented. The version 2.0 (beta) of correspSearch was presented at the TEI conference in Graz in 2019 and went into live production in summer 2021. The size of the dataset increased continuously in the last years - most of the contributions coming from the community of scholarly editions: there are now over 200,000 edited letters from over 360 publications indexed in correspSearch. This paper aims to outline the developments of the last years and give some insights to the current work, as well as an outlook on the upcoming version 3 of the web service.

Further developments in the last few years include e.g. new search functions based on information from authority files. The search options have been expanded by using not only the URLs from authority files for identification, but also the

information associated with the URIs to enrich the aggregated CMIF data. Of particular relevance in this regard is the German National Library's Integrated Authority File (Gemeinsame Normdatei)<sup>[1]</sup>, which in turn is supplemented with information from Wikidata<sup>[2]</sup>. This opens up new approaches to research, as for example a search for correspondences based on the gender or occupations of the correspondence partners.<sup>[3]</sup> In the late summer of 2023, new search functionalities based on extensions of CMIF v2 (Dumont et. al. 2019) will be released in a public beta phase, such as people or places mentioned in the letter.

Not only the search functionalities, but also the capture tools and workflows provided by correspSearch have been constantly developed further. For example, the CMIF Creator (Müller-Laackman 2022; Müller-Laackman, Dumont, Grabsch 2019) offers new functions to capture correspondence between (only) two correspondents more easily. The tools CMIF Check<sup>[4]</sup> and CMIF Preview<sup>[5]</sup> provide new ways to check the CMIF file for technical (and in parts also contentwise) consistency. In addition, correspSearch has also developed videos that provide an initial overview of the web service, as well as more detailed information on the digital methodological background. Another video tutorial guides the user step by step through the CMIF Creator to support the community in providing correspondence metadata of their edited letters.<sup>[6]</sup>

Finally, we would like to look ahead to developments, which will (probably) be still underway in late summer/fall. These include full text search with snippet view as well as visualizations that support the exploration of the dataset or individual search requests.

## Bibliography

Dumont, Stefan. 2016. „correspSearch – Connecting Scholarly Editions of Letters“. *Journal of the Text Encoding Initiative* (10). <https://doi.org/10.4000/jtei.1742>

Dumont, Stefan; Grabsch, Sascha; Müller-Laackman, Jonas (Eds.). 2022. *correspSearch – Connect Scholarly Editions of Correspondence (2.1.0)* [Webservice]. Berlin-Brandenburg Academy of Sciences and Humanities 2022. <https://correspSearch.net>

Dumont, Stefan, Ingo Börner, Dominik Leipold, Jonas Müller-Laackman, and Gerlinde Schneider. 2019. 'Correspondence Metadata Interchange Format'. In *Encoding Correspondence. A Manual for TEI-XML-Based Encoding of Letters and*

Postcards, edited by Stefan Dumont, Susanne Haaf, and Sabine Seifert. Berlin.  
<https://encoding-correspondence.bbaw.de/v1/CMIF.html>.

Müller-Laackman, Jonas, Dumont, Stefan, & Grabsch, Sascha. 2019. CMIF Creator – Digitale Briefverzeichnisse leicht erstellt. multimedial & multimodal (DHd 2019), Mainz / Frankfurt, Germany. Zenodo. <https://doi.org/10.5281/zenodo.2613462>

Müller-Laackman, Jonas. 2022. 'CMIF Creator - Creating a Digital Index of Letters'. Version 2.1.4. [Software] <https://doi.org/10.5281/zenodo.4075014>.

Stadler, Peter. 2012. „Normdateien in der Edition“. *editio* 26: 174–83.

Stadler, Peter. 2014. 'Interoperabilität von Digitalen Briefeditionen'. In *Fontanes Briefe Ediert*, edited by Hanna Delf von Wolzhagen, 278–87. Fontaneana 12. Würzburg: Königshausen & Neumann.

Stadler, Peter, Marcel Illetschko, and Sabine Seifert. 2016. 'Towards a Model for Encoding Correspondence in the TEI: Developing and Implementing <correspDesc>'. *Journal of the Text Encoding Initiative* [Online] 9. <https://dx.doi.org/10.4000/jtei.1742>.

TEI Correspondence SIG. 2018. *Correspondence Metadata Interchange Format* (CMIF). <https://github.com/TEI-Correspondence-SIG/CMIF>

## About the authors

**Stefan Dumont** has been a researcher at TELOTA/BBAW since 2011. He is also coordinator of the DFG project *correspSearch*, co-leader of the DFG project *The German Letter in 18th Century* and co-convener of the TEI SIG Correspondence. His research focuses on digital editions of correspondence.

**Sascha Grabsch** studied Literary Studies, Philosophy and Media Studies in Potsdam. Since 2012, he has been working at the BBAW in DFG projects and in the TELOTA initiative in the field of DH. The focus of his work and research is on the digital indexing of research data and scholarly editions of modern texts.

**Jonas Müller-Laackman** studied Arabic Studies in Berlin and Leiden. He was a researcher in the *correspSearch* Team between 2018 and 2021. Since 2022, he has been working as a Referent for Digital Research Services at the State- and University Library Hamburg. His work focuses on multilingual DH, DH methodologies and Arabic literature.

**Ruth Sander** studied English Philology, Scandinavian Studies, and Digital Humanities in Göttingen and Stuttgart. She joined TELOTA in 2021 where her primary focus is the digital indexing of research data and scholarly editions of modern texts.

**Steven Sobkowski** studied Media Informatics in Berlin and has been working as a Research Software Engineer since 2023 in TELOTA for the projects *correspSearch* and *Redaktions-/Onlinesystem für Online-Editionen des Bundesarchivs*.

## Notes

---

1. Cf. [https://www.dnb.de/EN/Professionell/Standardisierung/GND/gnd\\_node.html](https://www.dnb.de/EN/Professionell/Standardisierung/GND/gnd_node.html) ↩
2. Cf. [https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page) ↩
3. E.g. a search for correspondences from/to music publisher in the 19th century: <https://correspsearch.net/en/search.html?d=1800-1900&x=1&o=http://www.wikidata.org/entity/Q222836&w=0> ↩
4. E.g. <https://correspsearch.net/services/check/check.xql?url=https://correspsearch.net/storage/forster/Band13.xml> ↩
5. E.g. <https://correspsearch.net/services/cmif-preview.xql?url=https://correspsearch.net/storage/forster/Band13.xml> ↩
6. Cf. <https://correspsearch.net/de/videos.html> (german) and <https://correspsearch.net/en/videos.html> (english). More videos will be added over the next months. ↩

# Towards shared TEI model/s for institutional minutes and protocols – protokolleditionen.eu

Stephan Kurz, Austrian Academy of Sciences, Austria

Sven Jüngerkes, Kommission für Geschichte des Parlamentarismus und der politischen Parteien

Poster

**Keywords:** open governmental data, editing minutes, TEI as a social tool

## Abstract

A group of scholarly editors working on TEI editions of minutes/protocols is working on common best practices for editing this textual genre. The »Arbeitskreis Digitale Protokolleditionen« group underlines the collaborative and social aspect of TEI editing. Its proposed poster hence does not claim to present results but intends to showcase and encourage a collaborative culture of encoding.

Following a panel at the DHd2022 conference,<sup>[1]</sup> a group of scholarly editors working on TEI based editions of minutes (mostly from the political sphere at the moment) from the German speaking countries has met on a regular basis to exchange experiences, markup choices and common ground for exchanging data. In 2023, this loose group rebranded itself as *Arbeitskreis Digitale Protokolleditionen*.<sup>[2]</sup>

To date, the discussions within that group have targeted

- roles and functions of minutes/protocols
- TEI modelling decisions and formalizing of preexisting editing guidelines
- extraction of `tei:listEvent` data for defining an API for common calendar applications<sup>[3]</sup>
- strategies for reuse of prosopographical auxiliary/indexing data
- best practices and tools for manual, semi-automatic and automatised edition data enrichment
- sharing and reuse of bibliographical data, e.g. through Zotero

The poster aims to showcase the opportunities of collaboration across a variety of institutional backgrounds. It serves as an example that the intellectual infrastructure of a common vocabulary (=subset/s of the TEI guidelines) provides chances to discuss, among other issues,

- common ground between DH technologists and domain specialist editors
- possible ways of extracting overlaps in data
- while minimising data input duplication

These efforts are based on voluntary contribution of the *Arbeitskreis* members which relies on willingness for formal-informal potlatch style giving and taking, and on the institutional support of the democratic and academic institutions that fund digital editions of minutes. Apart from the concrete example that deals with retro-projecting ideas of open data into the past by unlocking (administrative) written fixations of institutional negotiations and decision making processes, we want to emphasize the inherent social impact of the TEI guidelines: They are a catalyst for discussing important questions of scholarly editing.

## Notes

---

1. <https://doi.org/10.5281/zenodo.6328072>. ↵
2. <https://www.protokolleditionen.eu/> ↵
3. This relates to ongoing TEI-C discussions on the model of `tei:event`, cf. <https://doi.org/10.5281/zenodo.3447297>; <https://github.com/TEIC/TEI/issues/2382> and related. ↵

# Transitional MerMEId – Responding to Community Needs in Software and Sustainability

**Margrethe Bue**, National Library of Norway, Oslo

**Jonathan Gammert**, Akademie der Wissenschaften und der Literatur Mainz, Germany

**Clemens Gubsch**, Austrian Academy of Sciences (ÖAW), Austria

**Daniel Jettka**, Universität Paderborn, Germany

**Severin Kolb**, SLUB Dresden

**David Lewis**, University of Oxford e-Research Centre

**Joshua Neumann**, Akademie der Wissenschaften und der Literatur Mainz, Germany

**Matthias Richter**, SLUB Dresden

**Kristina Richts-Matthaei**, Universität Paderborn, Germany

**Peter Stadler**, University of Paderborn, Germany

**Poster**

**Keywords:** metadata, digital work catalogues, sustainability, community driven software development

## Abstract

MerMEId is an established – and the only – tool in digital musicology that has been developed explicitly for creating and managing musical catalogue data in MEI. Its use in varied projects worldwide evinces its broad applicability. The MerMEId community regularly receives requests from musicological projects (catalogues, editions, etc.) that would like to use the tool for their specific needs and workflows. As such, a significant demand exists within the community for a stable and flexible MEI metadata editor. Originally developed as an in-house solution at the *Danish Center for Music Editing*, MerMEId has transitioned into a community driven, open-source project coordinated by *Virtueller Forschungsverbund Edirom* (ViFE) and the *Austrian Centre for Digital Humanities and Cultural Heritage* (ACDH-CH). While the change in governance – along with the establishment of platforms for regular communication and exchange – can be considered completed, the code basis of the software needs extensive revision for re-usability and sustainability.

Currently identified areas for updating include: more maintainable and modularized software, broad and nuanced testing (unit tests, regression tests, integration tests, end-to-end tests, performance tests), and better documentation for users and developers. Contextualized within an accounting of projects using the MerMEId and a review of its current functionality, this poster highlights these planned revisions and outlines a proposed infrastructure for enhancing sustainability. These changes aim to enhance usability and user-friendliness, thereby solidifying the MerMEId as the preeminent tool for creating digital catalogues.

These issues need substantial work that goes beyond the voluntary work of contributors from the community.



# Travelling Humboldt—Data on the Move

**Stefan Dumont**, Berlin-Brandenburg Academy of Sciences and Humanities,  
Germany

**Tobias Kraft**, Berlin-Brandenburg Academy of Sciences and Humanities, Germany

**Sabine Seifert**, Berlin-Brandenburg Academy of Sciences and Humanities,  
Germany

**Christian Thomas**, Berlin-Brandenburg Academy of Sciences and Humanities,  
Germany

**Jan Wierzoch**, Berlin-Brandenburg Academy of Sciences and Humanities, Germany

## Long Paper

**Keywords:** digital scholarly editions, single-source publishing, data publication, TEI Customization, authority files

The long-term Academy project *Travelling Humboldt—Science on the Move*, based at the Berlin-Brandenburg Academy of Sciences and Humanities in Germany, publishes the American, Russian-Siberian and European Travel Journals of the Prussian naturalist and explorer Alexander von Humboldt (1769–1859)<sup>[1]</sup>. The journals are accompanied by thematically related letters from his world-spanning correspondence network as well as manuscripts from his vast legacy collection, many of which have not been published before.

Six years after the first presentation at the *TEI Conference and Members' Meeting* 2016 in Vienna (Dumont et al. 2016), we have delivered eight subsequent versions of this digital, documentary edition<sup>[2]</sup>. The TEI/XML subset<sup>[3]</sup> of the *ehd* has been developed by adopting established TEI subsets, e.g. the German Text Archive's *Base Format for Manuscripts* (BBAW 2022; Thomas / Haaf 2017) and the CMIF format for correspondence metadata (TEI Correspondence SIG 2018), to ensure the highest possible degree of standardisation, re-usability, and interoperability of the data. The comprehensive transcription and encoding guidelines illustrate the specifications of the *ehd's* TEI format.<sup>[4]</sup> With the publication of the first volumes of its print component with the Springer Nature/J. B. Metzler publishing house<sup>[5]</sup>, the project's hybrid strategy has been realised. With an effective single source approach, both the digital and the print component (book, PDF and eBook derivatives) completely rely on the same TEI/XML encoded data. The publication

strategy is 'digital first', and the text-critical documentary edition can be fully accessed open access under a Creative Commons license.<sup>[6]</sup>

Following version 8 of the *ehd* (published in May 2022)<sup>[7]</sup>, we now fulfil the promise to deliver Humboldt's complex handwritten and hard-to-decipher texts *as data* by making available (1) the annotated text transcriptions of more than 500 documents (ca. 2.800 pages), (2) the comprehensive *Alexander von Humboldt Chronology* with about 1.600 individual, dated statements from Humboldt's almost 90-year lifespan, and (3) about 18.000 index entries (e.g. persons, places, institutions, bibliographic items) (Ette et al. 2022). All datasets are in TEI/XML format. The transcriptions and indices have been enriched using authority file information wherever available. Especially the index of persons shows great potential for re-use, since several historical (and mythological) persons in Humboldt's texts have not yet been documented in the most important authority file for the German-speaking research community, the GND of the German National Library<sup>[8]</sup>. Recently, we handed over a first batch of 50+ person index entries to the library which have already been integrated into the GND, with the *ehd* named as the authoritative source.<sup>[9]</sup> This supplementary data can help enhance the community's authority data within the GND, Wikidata and other international portals.

The presentation focuses on this digital data publication, and illustrates our approach between re-using existing data and best practices on the one hand, and giving back to the wider (TEI) community on the other hand. For example, the adopted formats and specifications that were enhanced to fulfil the document-specific encoding needs of Humboldt's writings can now be adopted by other projects. The *ehd* has also been active in discussions of the TEI recommendations, proposing specific interpretations or alterations of the guidelines. We demonstrate some of these developments with an emphasis on the most generally relevant aspects.<sup>[10]</sup>

## Bibliography

Annisius, Marie, Barbara Fischer, and Alexander Steckel. 2022. 'Ein Baum in Der GND - Gemeinsame Normdatei'. <https://doi.org/10.5281/zenodo.7245843>.

Berlin-Brandenburgische Akademie der Wissenschaften (BBAW) (Ed.). 2022. *DTABf. Deutsches Textarchiv – Basisformat*. <https://www.deutschestextarchiv.de/doku/basisformat/brief.html>

Kraft, Tobias; Dumont, Stefan. 2020.: "The Humboldt Code. On creating a hybrid digital scholarly edition of a 19th century globetrotter." In: *Wiener Digitale Revue* 1 (2020). DOI: <https://doi.org/10.25365/wdr-01-03-02> (all URLs in this abstract last accessed 2022-06-20).

Dumont, Stefan, with Susanne Haaf, Tobias Kraft, Alexander Czymiel, Christian Thomas, Matthias Boenig. 2016. "Applying Standard Formats and Tools: 'Alexander von Humboldt auf Reisen' as an Example for the Collective Subsequent Use of DTABf and ediarum". Presented at the *TEI Conference and Members' Meeting* 2016, Vienna, 2016, September 26–30. Abstract (PDF): [https://www.tei-c.org/Vault/MembersMeetings/2016/sites/default/files/TEIconf2016\\_BookOfAbstracts.pdf](https://www.tei-c.org/Vault/MembersMeetings/2016/sites/default/files/TEIconf2016_BookOfAbstracts.pdf), pp. 69–70.

Ette, Ottmar, Stefan Dumont, Annika Geiser, Carmen Götz, Tobias Kraft, Ulrike Leitner, Ulrich Päßler, Florian Schnee, and Christian Thomas. 2022. 'TEI-XML-Datensatz der Tagebücher, Briefe, Dokumente, Forschungsbeiträge, Chronologieeinträge und Register der edition humboldt digital'. [Datensatz] GitHub. <https://github.com/telota/edition-humboldt-digital>.

Fischer, Barbara. 2022. 'And the beat goes on - Die Zusammenarbeit von GND und Text+'. GND Confluence (blog). 6 September 2022. <https://wiki.dnb.de/pages/viewpage.action?pageId=253800499>.

Götz, Carmen, ed. 2022. Alexander von Humboldt: Tagebücher der Amerikanischen Reise: Von Spanien nach Cumaná. Vol. 1. edition humboldt print, I. Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-662-64274-0>.

Päßler, Ulrich, and Ottmar Ette, eds. 2020. Alexander von Humboldt: Geographie der Pflanzen: Unveröffentlichte Schriften aus dem Nachlass. Vol. 1. edition humboldt print, III. Stuttgart: J.B. Metzler. <https://doi.org/10.1007/978-3-476-04965-0>.

Pierazzo, Elena: "A Rationale of Digital Documentary editions". In: *Literary and Linguistic Computing*, 26,4 (2011), pp. 463–477. DOI: <https://doi.org/10.1093/lc/fqr033>.

Sahle, Patrick: "What is a Scholarly Digital Edition?" In: *Digital Scholarly Editing: Theories and Practices*. Edited by Matthew James Driscoll and Elena Pierazzo, Cambridge, UK: Open Book Publishers, 2016.

TEI Correspondence SIG. 2018. *Correspondence Metadata Interchange Format* (CMIF). <https://github.com/TEI-Correspondence-SIG/CMIF>

Thomas, Christian; Haaf, Susanne. 2017. "Enabling the Encoding of Manuscripts within the DTABf: Extension and Modularization of the Format". In: *Journal of the Text Encoding Initiative* [Online], Issue 10. DOI: <https://doi.org/10.4000/jtei.1650>.

## About the authors

**Stefan Dumont** has been a researcher at TELOTA/BBAW since 2011. He is also coordinator of the DFG project *correspSearch*, co-leader of the DFG project *The German Letter in 18th Century* and co-convener of the TEI SIG Correspondence. His research focuses on digital editions of correspondence.

**Tobias Kraft** works as a research coordinator at the BBAW. Together with his team, he creates the *edition humboldt*, a digital and printed edition of Alexander von Humboldt's travel manuscripts and personal papers. Since 2019, he is also the director of the *Proyecto Humboldt Digital* (ProHD).

**Sabine Seifert** works as a research associate in the project *A. v. Humboldt–Science on the Move* (BBAW) and at the Theodor Fontane Archive. She is co-convener of the TEI SIG Correspondence and TEI council member. Main interests are digital editions, manuscripts, history of humanities and sciences.

**Christian Thomas** has been working in the German Text Archive and the infrastructure project CLARIN. Currently he is focussing on hybrid scholarly editions of A. v. Humboldt's manuscripts & travel journals (BBAW), as well as Johann Wolfgang von Goethe's letters & diaries (Klassik Stiftung Weimar).

**Jan Wierzoch** has been a researcher at TELOTA/BBAW since 2020. He studied History as well as Library and Information Science. At TELOTA, he is involved in the development of several digital editions, with a focus on visualisations.

## Notes

- 
1. Project description: <https://www.bbaw.de/en/research/alexander-von-humboldt-auf-reisen-wissenschaft-aus-der-bewegung-travelling-humboldt-science-on-the-move>; digital edition: <https://edition-humboldt.de/>; cf. Kraft/Dumont 2020. ↩

2. On Digital Scholarly Editions see Sahle 2016; esp. on the concept of 'documentary editions' cf. Pierrazo 2011. [↵](#)
3. The ODD of the *ehd* ("chained" from DTABf and others) will be published in summer 2023. [↵](#)
4. <https://edition-humboldt.de/richtlinien/index.html> [↵](#)
5. Book series *edition humboldt print*, <https://www.springer.com/series/16345>: Päßler / Ette 2020; Götz 2022. [↵](#)
6. CC-BY-SA 4.0 (<https://creativecommons.org/licenses/by-sa/4.0/>) for the TEI/XML; CC-0 for indices and metadata. [↵](#)
7. Cf. the overview at <https://edition-humboldt.de/about/index.xql?id=H0020382&l=en>; API: <https://edition-humboldt.de/about/index.xql?id=api&l=en>; TEI/XML (current version of the *ehd*) <https://edition-humboldt.de/api/v1.1/tei-xml.xql>. [↵](#)
8. [https://www.dnb.de/EN/Professionell/Standardisierung/GND/gnd\\_node.html](https://www.dnb.de/EN/Professionell/Standardisierung/GND/gnd_node.html) [↵](#)
9. Cf., for instance, the entry on William Thomson, <https://d-nb.info/gnd/1283481235>. In the future, we plan to work together with GND even more efficiently via the Text+ GND Agency of the National Research Data Infrastructure/Text+, cf. Annisius, Fischer, Steckel 2022; Fischer 2022. [↵](#)
10. E.g. TEIC/TEI Issue #2028 "@calendar should allow multiple values", <https://github.com/TEIC/TEI/issues/2028>, following the discussion on the TEI mailing list in August 2020, leading to a corresponding implementation in the TEI P5 Guidelines v. 4.3.0 (2021-08-31). [↵](#)

# Updates and Outcomes from the Guatemalan Digitization and Encoding Project

Martha E. Thomae, McGill University, Canada

Julie Cumming, McGill University, Canada

Ichiro Fujinaga, Schulich School of Music, McGill University, Canada

Long Paper

**Keywords:** digitization and encoding workflow, white mensural notation, polyphonic choirbooks, machine-assisted transcription, Mensural MEI

## Abstract

In this paper, we are presenting the outcome of a four-year music-encoding project. Four years ago, at the Music Encoding Conference, we presented our plans for digitizing and encoding a mensural music corpus in Guatemala City. Six large sixteenth-century polyphonic choirbooks, copied during the Spanish colonial period, are held at Guatemala City's Metropolitan Cathedral. The plan presented consisted of a workflow for obtaining digital images and Mensural MEI files for these sources and said workflow was to be tested on one of the books (choirbook 1). The MEI files would encode the piece in score layout, allowing readers to appreciate the polyphonic texture of the music, which is obscured by the choirbook layout of the original sources where the voices are kept apart. Of course, the original workflow changed over four years, with some tools added and others replaced. In this paper, we present the updated workflow and the project outcomes to be used as a reference for other digitization and encoding projects, especially for those with limited resources.

The workflow has three steps: digitization, optical music recognition (OMR), and automatic voice alignment. Figure 1 shows these steps and the tools involved in each of them. The first step required building a do-it-yourself (DIY) book scanner, given that the archive lacked the appropriate digitization equipment for rare books. The other two encoding steps involved using free online software: MuRET and the Measuring Polyphony (MP) Editor.

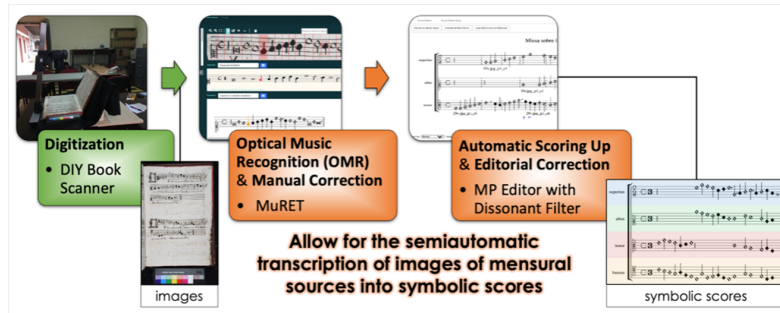


Fig. 1: Digitization and encoding workflow steps (in bold) and associated tools (in bullet points)

This paper is meant to provide the big picture of the workflow (with additional references for further details on its specific components), present updates and results (i.e., the digital images and Mensural MEI scores), and obtain feedback from the MEI/TEI community.

## Bibliography

Desmond, Karen, Andrew Hankinson, Laurent Pugin, Juliette Regimbal, Craig Sapp, and Martha E. Thomae. 2020. "Next Steps for Measuring Polyphony: A Prototype Editor for Encoding Mensural Music." In *Music Encoding Conference Proceedings 26–29 May, 2020*, 121–24. Tufts University, Boston. <http://dx.doi.org/10.17613/5k88-9z02>.

Desmond, Karen, Laurent Pugin, Juliette Regimbal, David Rizo, Craig Sapp, and Martha E. Thomae. 2021. "Encoding Polyphony from Medieval Manuscripts Notated in Mensural Notation." In *Proceedings of the Music Encoding Conference*, edited by Stefan Münnich and David Rizo, 197–219. Alicante, Spain (online): Humanities Commons. <https://doi.org/10.17613/tf2j-x697>.

Morales Abril, Omar. 2015. "Música local y música foránea en los libros de polifonía de las Catedrales de Guatemala y México." *Jahrbuch für Renaissance-musik: troja. Vokalpolyphonie zwischen Alter und Neuer Welt: Musikalische Austauschprozesse zwischen Europa und Lateinamerika im 16. und 17. Jahrhundert* (14): 95–124.

Pujol, David. 1965. "Polifonía española desconocida conservada en el Archivo Capitular de la Catedral de Guatemala y de la Iglesia parroquial de Santa Eulalia de Jacaltenango." *Anuario Musical* 20: 3–10.

Rizo, David, Jorge Calvo-Zaragoza, and José M. Iñesta. 2018. "MuRET: A Music Recognition, Encoding, and Transcription Tool." In *Proceedings of the 5th International Conference on Digital Libraries for Musicology (DLfM)*, 52–56. Paris, France: ACM. <https://doi.org/10.1145/3273024.3273029>.

Snow, Robert J., ed. 1996. *A New-World Collection of Polyphony for Holy Week and the Salve Service: Guatemala City, Cathedral Archive, Music MS 4*. Monuments of Renaissance Music 9. Chicago, IL: University of Chicago Press.

Stevenson, Robert. 1970. *Renaissance and Baroque Musical Sources in the Americas*. Washington, DC: General Secretariat, Organization of American States.

Thomae, Martha E., Julie E. Cumming, and Ichiro Fujinaga. 2022a. "Digitization of Choirbooks in Guatemala." In *Proceedings of the 9th International Conference on Digital Libraries for Musicologists (DLfM)*, 19–26. Prague, Czech Republic: ACM. <https://doi.org/10.1145/3543882.3543885>.

Thomae, Martha E., Julie E. Cumming, and Ichiro Fujinaga. 2022b. "Counterpoint Error-Detection Tools for Optical Music Recognition of Renaissance Polyphonic Music." In *Proceedings of the 23rd International Society for Music Information Retrieval Conference (ISMIR)*. Bengaluru, India.

## About the author

**Martha Thomae** is a researcher at the DDMAL Lab at McGill University. She has a BS in Mathematics from Universidad del Valle de Guatemala and a Master's and PhD in Music Technology from McGill. Her research interests include music information retrieval, early music, and cultural heritage preservation. She is working on digitizing and encoding a Guatemalan polyphonic choirbook using music-encoding technologies.



# Using TEI for Chinese Architectural Data

**Tracy Miller**, Vanderbilt University, United States of America

**Clifford Anderson**, Vanderbilt University, United States of America

**Jing Zhuge**, Southeast University, Nanjing, PR China

**Lala Zuo**, New York University, Shanghai, PRC, China

**Aurelia Campbell**, Boston College, United States of America

**Short Paper**

**Keywords:** TEI, IIIF, Digital Chinese  
Architecture, Cultural Heritage  
Preservation

## Abstract

Architectura Sinica (<https://www.architecturasinica.org/>) is a browser-based research portal dedicated to the study of traditional Chinese Architecture. The open-source database is designed to facilitate advanced research on Chinese linguistic. Architectura Sinica uses TEI as its single source of truth, relying on XQuery for search and discovery to enhance trans-disciplinary research on the built environment of pre-modern China.

Building on the Srophé App (<https://srophe.app/>), an open-source eXistDB application for digital cultural heritage, Architectura Sinica is customized for research on historic architecture. Analysis of Middle Period (900-1200 CE) Chinese timber-frame architecture requires detailed investigation of the complex structures supporting the weight of ceramic roof tiles and its characteristic mortise and tenon bracketing. To foster understanding of these intricate structures, connecting TEI-based metadata with images is crucial.

Currently, Architectura Sinica uses Flickr/SmugMug to store its image archive of Chinese temple buildings and illustrations of technical terminology. Although a good pragmatic choice initially, reliance on this commercial provider increasingly poses technical barriers to its scholarly ends. By connecting the TEI with IIIF, we expect to resolve three pressing issues with our project: stronger visual analysis in the interface, better search and discovery of the images, and increased integration of image metadata in the TEI.

This talk will introduce Architectura Sinica, focusing on how it developed from the Srophé app to support an international community of scholars active in the field of medieval Chinese religious architecture. We discuss how the combination of TEI, XQuery, and eXistDB have simplified our application architecture and how these technologies fostered our transition to IIIF. By documenting this phase of Architectura Sinica, we hope to encourage other scholars to consider using TEI and related XML-related technologies to document and preserve cultural heritage in their domains.

## About the authors

**Tracy Miller** is Associate Professor in the Department of History of Art and Architecture and Department of Asian Studies at Vanderbilt University. She holds an interdisciplinary MA and PhD in Asian Studies/Art-Architecture History from the University of Pennsylvania (2000) and has published *The Divine Nature of Power: Chinese Ritual Architecture at the Sacred Site of Jinci* (Harvard Asia Center, 2007), as well as articles and chapters on the religious architecture of Medieval China (ca. 500-1200) in *Art Bulletin*, *Asia Major*, *Archives of Asian Art*, *Artibus Asiae*, and other venues. An active member of Vanderbilt's Digital Humanities community, she has spearheaded the development of ArchitecturaSinica.org, a public, collaborative, web-based research portal for the study of Traditional Chinese Architecture, as part of a long-term project analyzing the influence of religious goals on regional style in China's earliest extant timber-frame structures.

**Clifford B. Anderson** is Director of Digital Research at the Center of Theological Inquiry in Princeton, NJ and Chief Digital Strategist at the Vanderbilt University Library. He holds a secondary appointment as Professor of Religious Studies in the College of Arts & Science at Vanderbilt University. From 2018 to 2020, he was also an Adjunct Professor of Computer Science in the Department of Electrical Engineering and Computer Science in the Vanderbilt University School of Engineering. Anderson has a M.Div. from Harvard Divinity School, a Ph.D. from Princeton Theological Seminary, and a M.S. in L.I.S. from the Pratt Institute. Among other works, he is co-author of *XQuery for Humanists* (Texas A&M Press, 2020) and editor of *Digital Humanities and Libraries and Archives in Religious Studies*.

# Valid, but readable: writing D2d for constructing XML

Markus Lepper, semantics gGmbH, Germany

Baltasar Trancón y Widemann, semantics gGmbH, Germany; Nordakademie Elmshorn, Germany

Long Paper

**Keywords:** Lenient parsing, markup minimization, man-machine-interaction, creative flow

## Abstract

D2d is a text format definition which enables technical authors, journalists, scholars, poets, essayists, novelists, and other non-it-experts, to notate a structured text in the creative flow, without intervening disruptive technical actions, or even without any technical devices at all. Nevertheless the notated text can represent an XML encoded document valid w.r.t. a particular document type. (Lepper et.al 2001) (Lepper-Trancón 2016, 2019)

D2d employs only two redefinable escape characters, for commands and comments. Therefore any d2d encoded document is easily readable and writable by computers and humans. It can be notated, with the same level of precision, using arbitrary text editors, pen and pencil, chalk on a blackboard. It can be communicated by speech, sign language and stenography. Its input format is inspired by m4 (Kernighan-Ritchie 1977), LaTeX (Lamport 1986), and Lout (Kingston 1992).

With its compiler implementation, efficient and comfortable work-flow has been established in very different areas of application, like musicology (Lepper 2015), interactive HTML pages, accounting, relational data base input, technical configuration files, etc. (Lepper-Trancón 2016)

Nearly-WYSIWYG editors like Oxygen (Oxygen 2023) are excellent tools when one particular document type shall be edited with one particular rendering, and thus the XML as such can be made (almost) transparent to the domain expert. They perform less when the XML structure is itself the subject of editing, like in most contexts of TEI. Naturally, they do not cover the other media listed above. XML tags, escapes, entities, etc. have been designed for data transmission. Typing them manually

requires either frequent technical intervention of the editing software or many redundant key strokes. Both can disturb or even destroy the creative flow of writing.

The d2d compiler is an alternative, which applies to plain text files advanced techniques on two levels of parsing: The upper uses explicit tagging. Nearly all closing tags and some opening tags are unambiguously inferred. The lower level infers all tags from character-oriented parser definitions, suitable for small complex data like embedded mathematical or chemical formulas, calendric dates, sigla, abbreviations, etc. Both levels directly translate to valid XML.

In the discussion of Diachronic Markup, Barney (2018) gives as an example a list of change activities of the poet, in a sequential order proposed by the scholar. In canonical XML encoding this looks like ...

```
<listChange>
  <change xml:id="ch01">
    <desc>writes "I am blac"</desc>
  </change>
  <change xml:id="ch02">
    <desc>overstrikes "blac"</desc>
  </change>
  <change xml:id="ch03">
    <desc>writes "a curse:"</desc>
  </change>
  <-- ... -->!
  <change xml:id="ch07">
    <desc>writes "a black slave spoke me"</desc>
  </change>
  <-- ... -->!
  <change xml:id="ch09">
    <desc>overstrikes "black slave" and writes
      "negro"</desc>
  </change>
  <change xml:id="ch10">
    <desc>overstrikes "spoke" and writes "felt"</desc>
  </change>
  <change xml:id="ch11">
    <desc>overstrikes "felt" and writes "thought"</desc>
  </change>
  <change xml:id="ch12">
    <desc>overstrikes "thought" and writes "thinks"</desc>
  </change>
</listChange>
```

Listing 1

An equivalent d2d file is ...

```
#d2d module changes
  public tags changelist = change*
  tags change = #implicit id, (overstrikes, writes?) | writes
  tags id, overstrikes, writes = #char
    with xmrep trimmed
end module

#d2d 2.0 text using changes:changelist
  #change ch01 #writes I am blac
  #change ch02 #overstrikes blac
  #change ch03 #writes a curse:
  #change ch07 #writes a black slave spoke to me
  #change ch09 #overstrikes black slave #writes negro
  #change ch10 #overstrikes spoke      #writes felt
  #change ch11 #overstrikes felt      #writes thought
  #change ch12 #overstrikes thought   #writes thinks
#eof
```

Listing 2

The author's "For each <change> element I then added a <desc> containing human-readable prose" is made obsolete by mark-up which is immediately both: readable by humans and by computer.

In that article follows a rendering of the poem with all corrections. In d2d this can look like ...

```
#d2d module changedText
  public tags text = (#chars | del)*
  tags del = @text, rep*
  tags rep = #implicit header, @text
  chars header = ([overstrike: "-"]|[crossout: "x"])?, [id: ('0'..'9')+]?
end module

#d2d 2.0 text using changedText:text
I am #del(blac #rep -01) a curse: a
      #del(black slave
      #rep -09 negro) #del(spoke
          #rep -10 felt
          #rep -11 thought
          #rep -12 thinks) to me
#eof
```

Listing 3

The slimness of d2d tags allows to arrange mark-up geometrically similar to the manuscript, to reflect deeply nested editing:

```
The #del(#del<pupil
  #rep young #del( man
    #rep { person)
  #rep x reader> must be aware
  #rep x public should know) that the following lines are meant ...
```

#### Listing 4

Indentation is for humans only. The computer executes normal left-to-right parsing; corrections appearing “above” in the manuscripts are now “below”. Nevertheless the readability is much better than genuine XML encoding, into which the data will be compiled for further processing:

```
The <del><del>pupil
    <rep>young <del>man
        <rep overstrike="{\}> person</rep>
        </del>
    </rep>
    <rep overstrike=x">reader</rep> must be aware
</del>
    <rep overstrike=x">public should know</rep>
</del>that the following lines are meant ...
```

#### Listing 5

(Both approaches currently lack a formal definition of their semantics, which must precede any comparison!)

In the first two d2d examples, the few lines at the beginning give the *complete necessary* document type definition: Both files are totally self-contained.

Both parsing levels employ advanced techniques for generating error messages to guide domain experts (non-computer-language-experts) through the intricacies of grammar definitions, and can accept partially invalid input for rapid prototyping, generating output with embedded error messages. (Lepper-Trancón 2011)

The compiler delivers standard XML representations: file content or SAX event stream. It can run directly on a DTD, or can use its proprietary type definition format. This employs *rewriting for parametrization*, a new technique allowing light-weight derivations of new document types. (Lepper-Trancón 2018)

A special mode supports XSLT program code for a given XML backend format, including approximate, light-weight type checking. (Lepper-Trancón 2015)

## Bibliography

Barney, Brett (2018) *TEI, the Walt Whitman Archive, and the Test of Time.*, in: Journal of the Text Encoding Initiative, 13, Selected Papers from the 2018 TEI Conference. <https://doi.org/10.4000/jtei.3249>

Kernighan, Brian W. and Ritchie, Dennis M. (1977) *The M4 macro processor. Technical report*, Bell Laboratories, Murray Hill (NJ).

Kingston, Jeffrey H. (1992) *The Design and Implementation of the Lout Document Formatting Language*, in: *Software---Practice & Experience* 23 (9).  
<http://citeseer.ist.psu.edu/kingston93design.html>

Lepper, Markus and Trancón y Widemann, Baltasar and Wieland, Jacob (2001) *Minimize Mark-Up! - Natural Writing Should Guide the Design of Textual Modeling Frontends*, in: *Conceptual Modeling -- ER2001, LNCS, Vol. 2224*, Springer, Berlin, November 2001.

Lepper, Markus and Trancón y Widemann, Baltasar (2011) *D2d --- a Robust Front-End for Prototyping, Authoring and Maintaining XML Encoded Documents by Domain Experts*, in: *Proceedings of KEOD 2011 International Conference on Knowledge Engineering and Ontology Design*, SciTePress, Portugal, 2011, p. 449-456.

Lepper, Markus and Trancón y Widemann, Baltasar (2015) *A Simple and Efficient Step Towards Type-Correct XSLT Transformations*, in: *Proceedings 26th International Conference on Rewriting Techniques and Applications (RTA 2015)*, LIPICS, Vol. 36, Dagstuhl Publishing, Saarbrücken/Wadern, 2015, p. 350-364.

Lepper, Markus (2015) *Gustav Mahler, Dritte Sinfonie, Erste Abtheilung. Eine Annäherung* [http://senzatempo.de/mahler/gmahler\\_sinf3\\_satz1.html](http://senzatempo.de/mahler/gmahler_sinf3_satz1.html)

Lepper, Markus and Trancón y Widemann, Baltasar (2016) *D2d - Kreatives Schreiben von XML-codierten Texten*, in: *Informatik 2016 (GI Jahrestagung)*, ed. by C. Mayr, Martin Pinzge, p. 1935-1940.

Lepper, Markus and Trancón y Widemann, Baltasar (2018) *Rewriting for Parametrization*, in: *Tagungsband des 35ten Jahrestreffens der GI-Fachgruppe "Programmiersprachen und Rechenkonzepte"*, IFI Reports, University of Oslo, Oslo, 2018.

Lepper, Markus and Trancón y Widemann, Baltasar (2019) *D2d -- XML for Authors*, in: *Tagungsband des 36ten Jahrestreffens der GI-Fachgruppe "Programmiersprachen und Rechenkonzepte"*, IFI Reports, University of Oslo, Oslo, 2019.

Lamport, Leslie (1986) *LaTeX User's Guide and Document Reference Manual*, Addison-Wesley, Reading (MA).

Oxygen (2023) Official oXygen website: <https://www.oxygenxml.com>

## About the authors

**Markus Lepper**, composer, music theorist, and computer scientist. He holds a PhD each in computer science and musicology. Works and lives in Berlin and is co-founder of semantics gGmbH Berlin.

**Baltasar Trancón Widemann** holds a PhD in computer science from TU Berlin and a Habilitation degree from the University of Bayreuth. He has worked as a researcher in academia and as an industrial software engineer. He is currently professor of programming at Nordakademie Elmshorn, and co-founder of semantics gGmbH Berlin.



# Visual or Symbolic? Best Practices for Encoding Neumes

**Elsa De Luca**, Nova University Lisbon, Portugal

**Inga Behrendt**, University for Church Music of the Diocese of Rottenburg-Stuttgart, Germany

**Ichiro Fujinaga**, Schulich School of Music, McGill University, Canada

**Kate Helsen**, Don Wright Faculty of Music, University of Western Ontario, Canada

**Stefan Morent**, Tübingen University, Germany

**Long Paper**

**Keywords:** MEI Neumes Module, Plainchant, St Gall notation, Symbolic versus Iconic

## Abstract

A general standard practice in MEI music encoding is to capture the meaning of the music symbols rather than their graphic appearance. While this philosophy is suitable for modern notations, it does not necessarily apply to the encoding of neumes. Early notations (9th – 13th century) conveyed a different set of musical instructions from what we are used to seeing in modern notation because it existed primarily in an oral tradition.

The original meaning of early music scripts is, to various degrees, lost. Early music palaeographers (Atkinson, Treitler, Rankin, among others) managed to pin down some principles behind the various music scripts employed across Europe and part of the Middle East and the Levant. Broadly speaking, the meaning originally attributed to the neumes depended on their graphical appearance (e.g., a vertical stroke mirroring a raising melody) and/or on some conventions established and shared by the scribes and readers familiar with that specific music script. The quantity of musical information conveyed by the neumes to modern readers varies according to the style of the music script. As such, the music encoding of early music scripts requires, on one hand, a system flexible enough to capture musical information unclear to the modern reader (e.g., a rising melody, even though the size of the interval is unknown). On the other hand, it should be able to express something (e.g., a neume) whose original meaning is now lost and the only thing left is a recognizable pen-stroke found repeatedly in the sources.

Recently some historical musicologists proposed some changes to improve the MEI Neumes Module applicability to different kinds of early notations. One of the specific challenges they faced was the difficulty of encoding potentially meaningful visual aspects of the neumes in the current MEI system without being certain of the semantic–musical meaning. Existing MEI schemas will need to be expanded to include several new attributes and elements in order to properly capture the rich graphical variety found in early music scripts.

This paper proposal aims to bridge the gap between semantic and graphic uses of MEI. We will present a case study highlighting the decision-making process behind some of the changes that will be suggested in the near future to the current MEI Neumes Module. By discussing some neumes in St Gall notation, we aim to contribute to the wider discussion about best practices for the encoding of neumes and, more broadly, to the digital representation of music. Hopefully this paper will give an insight on the laborious process, and specific expertise, required to encode neumes and contribute towards the conversation about the encoding of non-Western music notations. We wish to raise awareness on the peculiarity of neumatic notations within the bigger picture of musical scores and to foster dialogue within the MEI community and beyond, to find together the best system to encode neumes.

## Bibliography

Bain, Jennifer, Inga Behrendt, and Kate Helsen. 2014. "Linienlose Neumen und ihre Repräsentation mit MEI Schema, Herausforderungen in der Arbeit im Optical Neume Recognition Project (ONRP)." *Digitale Rekonstruktionen mittelalterlicher Bibliotheken*. Edited by Sabine Philippi and Philipp Vanscheidt. Trierer Beiträge zu den historischen Kulturwissenschaften 12: 119–132.

Behrendt, Inga, Jennifer Bain, and Kate Helsen. 2017. "MEI Kodierung der frühesten Notation in linienlosen Neumen." *Kodikologie und Paläographie im Digitalen Zeitalter 4 / Codicology and Palaeography in the Digital Age*. Vol. 4. Edited by Hannah Busch, Franz Fischer, and Patrick Sahle, with the cooperation of Philip Hegel and Celiz Krause, Norderstedt 2016. Köln: Institut für Dokumentologie und Editorik e.V, 2017, 281–296.

De Luca, Elsa, and Haig Utidjian. 2023. "Challenging the MEI Neumes Module: Encoding Armenian Neumes." In *Proceedings of the Music Encoding Conference*,

Dalhousie University, 19–22 May 2022, ed. by Jennifer Bain and David M. Weigl. Forthcoming.

De Luca, Elsa, Jennifer Bain, Inga Behrendt, Ichiro Fujinaga, Kate Helsen, Alessandra Ignesti, Debra Lacoste, Sarah Ann Long. 2019. "Capturing Early Notations in MEI: The Case of Old Hispanic Neumes." *MusikTheorie-Zeitschrift für Musikwissenschaft* 3: 229–249.

De Luca, Elsa, Jennifer Bain, Inga Behrendt, Ichiro Fujinaga, Kate Helsen, Alessandra Ignesti, Debra Lacoste, and Sarah Ann Long. "Cantus Ultimus' MEI Neume Module and its Interoperability Across Chant Notations." Abstract. *Music Encoding Conference*, Vienna. Available at [https://music-encoding.org/conference/abstracts/abstracts\\_mec2019/MEI%202019\\_De%20Luca%20et%20alii](https://music-encoding.org/conference/abstracts/abstracts_mec2019/MEI%202019_De%20Luca%20et%20alii).

Helsen, Kate, Inga Behrendt, and Jennifer Bain. 2017. "A Morphology of Medieval Notations in the Optical Neume Recognition Project." *Arti musices: Croatian Musicological Review* 48/2: 241–266. [https://hrcak.srce.hr/index.php?show=clanak&id\\_clanak\\_jezik=284211](https://hrcak.srce.hr/index.php?show=clanak&id_clanak_jezik=284211)

Hoppe, Paul, and Stefan Morent. 2021. "Computergestützte Tools zur Codierung des Gregorianischen Chorals. Ein neuer Eingabe-Editor für das MEI neumes-module." *Schaffen und Nachahmen. Kreative Prozesse im Mittelalter*, hrsg. v. Volker Leppin unter Mitarbeit von Samuel J. Raiser (Das Mittelalter. Perspektiven mediävistischer Forschung. Beihefte 16), Berlin und Boston 2021, 555–560.

Morent, Stefan, Fabian Köninger, and Niels Pfefffer 2023. *MEI Neumes Workspace*. <https://www.digimuwi.uni-tuebingen.de/neumes-workspace/>

Morent, Stefan. 2011. "Digitalisierungskonzepte für Neumen-Notationen - die Projekte Tübingen und e-sequence." *Perspektiven Digitaler Musikedition. Die Tonkunst* 3: 277–283.

## About the authors

**Elsa De Luca** is an early music scholar pursuing research on medieval notations at the Centre for the Study of the Sociology and Aesthetics of Music, NOVA University Lisbon. She is the PI of the interdisciplinary project *Echoes from the Past: Unveiling a Lost Soundscape with Digital Analysis*.

**Inga Behrendt** is professor of plainchant and hymnology at the University for Church Music of the Diocese of Rottenburg-Stuttgart and also teaches both

subjects at the University of Music, Freiburg. Her research concentrates on plainchant, digital plainchant research, and music historiography.

**Ichiro Fujinaga** is Canada Research Chair in Music Information Retrieval and Professor in the Music Technology Area of the Department of Music Research at the Schulich School of Music, McGill University. He has been working on encoding neume notation into MEI since 2010.

**Kate Helsen** is Assistant Professor in the Department of Music Research and Composition at the Don Wright Faculty of Music at the University of Western Ontario. Her specialization in plainchant, and Early Music more broadly, has led to her involvement in several international projects that dismantle the traditional boundaries of music and technology.

**Stefan Morent**, Professor/chair digital musicology/music before 1600 at Tübingen University, speaker of the Exploration Full Fund *Sacred Sound*. Research: Music of the Middle Ages, Performance Practice, Digital Musicology. International concerts with his ensemble *Ordo Virtutum* for Medieval Music.



