



TEI CONFERENCE AND MEMBERS' MEETING 2016 BOOK OF ABSTRACTS

TEI Conference and Members' Meeting 2016. Book of Abstracts

#TEIconf2016 – <http://tei2016.acdh.oeaw.ac.at>

Sponsors



ISBN: 978-3-200-04689-4

Edited by Claudia Resch
Vanessa Hannesschläger
Tanja Wissik

Austrian Centre for Digital Humanities
Austrian Academy of Sciences, Vienna



TEI CONFERENCE AND MEMBERS' MEETING 2016

September 26-30, Vienna

BOOK OF ABSTRACTS

Edited by Claudia Resch
Vanessa Hannessschläger
Tanja Wissik

Austrian Centre for Digital Humanities
Austrian Academy of Sciences, Vienna

TEI Conference and Members' Meeting 2016. Book of Abstracts

Edited by Claudia Resch, Vanessa Hanneschläger, and Tanja Wissik

Layout & Graphic Design

Sandra Lehecka

Austrian Academy of Sciences
Austrian Centre for Digital Humanities
Sonnenfelsgasse 19, 1010 Vienna, Austria
www.oeaw.ac.at/acdh

druck.at
Druck- und Handelsgesellschaft mbH
2544 Leobersdorf, Austria
www.druck.at

This book is licensed under a
[Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/) (CC-BY 4.0)



The information and views set out in this publication are those of the authors and do not necessarily reflect the opinion of the editors.

Vienna, September 2016
Printed in Austria
ISBN 978-3-200-04689-4

TEI 2016 Committees

Program committee **Martin Mueller** (chair)
Vanessa Hannessschläger
Serge Heiden
Martin Holmes
Stefan Majewski
Elli Mylonas
Claudia Resch
Martina Scholger
Daniel Schopper
Magdalena Turska
Tanja Wissik

Local committee **Claudia Resch** (chair)
Vanessa Hannessschläger
Sandra Lehecka
Daniel Schopper
Andrea Steiner
Tanja Wissik



Host & Venue

**The Austrian Centre for Digital Humanities
of the Austrian Academy of Sciences**





The Austrian Centre for Digital Humanities of the Austrian Academy of Sciences

The Austrian Centre for Digital Humanities (ACDH) of the Austrian Academy of Sciences (OeAW) is proud to host the TEI Conference and Members' Meeting 2016. The ACDH is an institute founded in 2014, pursuing research in various fields of the digital humanities and offering services for digitally working humanities scholars. Fundamental principles behind the ACDH's agenda include the systematic use of digital infrastructures, transdisciplinary, collaborative work and participatory technologies, open access / open source as well as open life cycles of research results and research data. The ACDH puts a strong focus on the technical support of digital editions and dictionaries working with TEI.

The Austrian Academy of Sciences is the leading Austrian non-university institution for science and research. It stands for social discourse, the transfer of knowledge and basic research at the highest international level. It is located in Vienna's first district in the heart of the city.



Photo: Sonnenfelsgasse 19, 1010 Vienna © Sandra Lehecka, ACDH-OeAW



Table of Contents

Introduction	19
Keynote	
<i>Tara Lee Andrews</i> Freeing our Texts from their (Digital Tool) chains	23
Plenary	
<i>TEI Technical Council</i> Whither TEI: The Next Thirty Years	27
Papers	
<i>Florentina Armaselu</i> Writing under the magnifying glass. Encoding the text in progress	31
<i>Piotr Bański, Bertrand Gaiffe, Patrice Lopez, Simon Meoni, Laurent Romary, Thomas Schmidt, Peter Stadler, Andreas Witt</i> Wake up, standOff!	35
<i>Michael Beißwenger, Eric Ehrhardt, Axel Herold, Harald Längen, Angelika Storrer</i> Converting and Representing Social Media Corpora into TEI: Schema and best practices from CLARIN-D	39
<i>Helena Bermudez Sabel</i> Tomayto, tomahto? Encoding variant taxonomies in TEI	43
<i>Alejandro Bia</i> TEI document and schema visualization and modeling using mindmaps	45
<i>Roman Bleier, Richard Breen</i> Documenting Transmission: The Analysis of the Folk Process using Versioning Machine 5.0	49
<i>Roman Bleier, Richard Hadden</i> Capturing the crowd-sourcing process: storing different stages of crowd-sourced transcriptions in TEI	51

<i>Jack Thompson Bowers, Thierry Declerck</i> TEI and lemon: a comparative study on the lexical encoding and interoperability	53
<i>Jack Thompson Bowers, Melanie Seltmann</i> Exploring data models for heterogenous dialect data: the case of explore.bread.AT!	55
<i>William Michael Broughton</i> Encoding Text as a Nexus of Perspective Hierarchies: Case Studies in XML, LMNL, and RDF	57
<i>Fabio Ciotti</i> We are TEI. You will be assimilated!	59
<i>James Cummings</i> Purifying and Simplifying: Advanced TEI Customization	63
<i>James Cummings, Jonathan Prag, James Chartrand</i> Upconversion and Migration: Generating a TEI-EpiDoc Corpus of Sicilian Inscriptions	65
<i>Michelle Dalmau, Meridith Beck Mink, James Voelkel, John A. Walsh, Wallace Hooper, William R. Newman, Archie Fields III</i> The Authoritative Bibliography for the Chymistry of Isaac Newton	67
<i>Stefan Dumont, Susanne Haaf, Tobias Kraft, Alexander Czmieł, Christian Thomas, Matthias Boenig</i> Applying Standard Formats and Tools: "Alexander von Humboldt auf Reisen" as an Example for the Collective Subsequent Use of DTABf and ediarum	69
<i>Yasmin Faghihi, Tobias J. Jocham</i> Islamic Manuscripts in Cambridge Digital Library & the German-French virtual library project Paleocoran	71
<i>Elena González-Blanco, Antonio Robles-Gómez, Salvador Ros, Gimena Del Rio Riande, Roberto Hernández, Miguel Urizar, Clara I. Martínez, Rafael Pastor, Jesús Cano, Agustín C. Caminero</i> TEIScribe: A graphic tool for composing and testing TEI documents in the context of the EVI-LINHD environment	73

<i>Vanessa Hanneschläger</i> Common Creativity international. CC-licensing and other options for TEI-based digital editions in an international context	77
<i>Matthew Lawrence Holford, Huw Jones, James Cummings</i> msDesc at Oxford and Cambridge	79
<i>Robert Klugseder, Christian Steiner</i> Cantus Network – a semantically enriched digital edition of libri ordinarii of the Salzburg metropolitan province	81
<i>Judit Kodolányi</i> Móricz Digital Edition – Possibilities and connections of a correspondence material	83
<i>Patrick McAllister</i> Does this shoe fit? Applying the TEI guidelines to Sanskrit philosophical texts	85
<i>Karlheinz Mörth, Daniel Schopper</i> Wiki2TEI: Wikipedias as sources for language research	87
<i>Anja Morgenstern, Agnes Amminger</i> Biography as compilation: How to edit Georg Nikolaus Nissen's "Biographie W. A. Mozart's" (1828) in TEI	89
<i>Matija Ogrin, Tomaž Erjavec</i> Early Modern Slovenian Manuscripts Between Description, Critical Edition and Lexicon	91
<i>Brian L. Pytlik Zillig</i> Shakespeare's N-grams at 400: Comparing Content Themes With Four Contemporaries	93
<i>Claudia Resch, Daniel Schopper</i> Encoding crime and punishment: A digital collection of historic broadsheets	95
<i>Chris Rogers</i> Creating a maintainable cataloging workflow and infrastructure using TEI	97

<i>Daniel Schopper, Kira Kovalenko, Thierry Declerck, Eveline Wandl-Vogt</i> Encoding a Dictionary of Russian Dialects in TEI and linking to LOD Resources	99
<i>Daniela Monika Schulz, Nils Geißler, Marcello Perathoner</i> Building a collaborative editorial workbench for legal texts with complex structures	101
<i>Martin Sievers</i> T ³ – Typesetting TEI Using TeX. A Plea for High-quality Typesetting in the Humanities	103
<i>Daniil Skorinkin, Evgeny Mozhaev</i> TEI markup for the 90-volume edition of Leo Tolstoy's complete works	107
<i>Magdalena Turska, Wolfgang Meier</i> TEI Processing Model Toolbox	109
<i>Toshinori Uetani, Sandrine Breuil, Mathieu Duboc, Guillaume Porte</i> The BVH in Tours and the "Bibliothèques françaises" project	111
<i>Georg Vogeler</i> Encoding Text about Things	113
<i>Dennis Zielke, Alexander Petrus</i> Automating the validation of TEI metadata processing	117

Posters

<i>Peter Andorfer, Dario Kampkaspar</i> HowTo create your own digital edition web app – A blog	121
<i>André De Tienne</i> STEP Transcriptor and STEP Text Comparator	123
<i>Thierry Declerck, Karlheinz Mörth</i> Towards a Repository of Senses for Use in TEI encoded Dictionaries	125
<i>Greta Franzini, Peter Andorfer</i> The Digital Editions Catalogue App: A web application to browse, curate and analyze digital editions	127

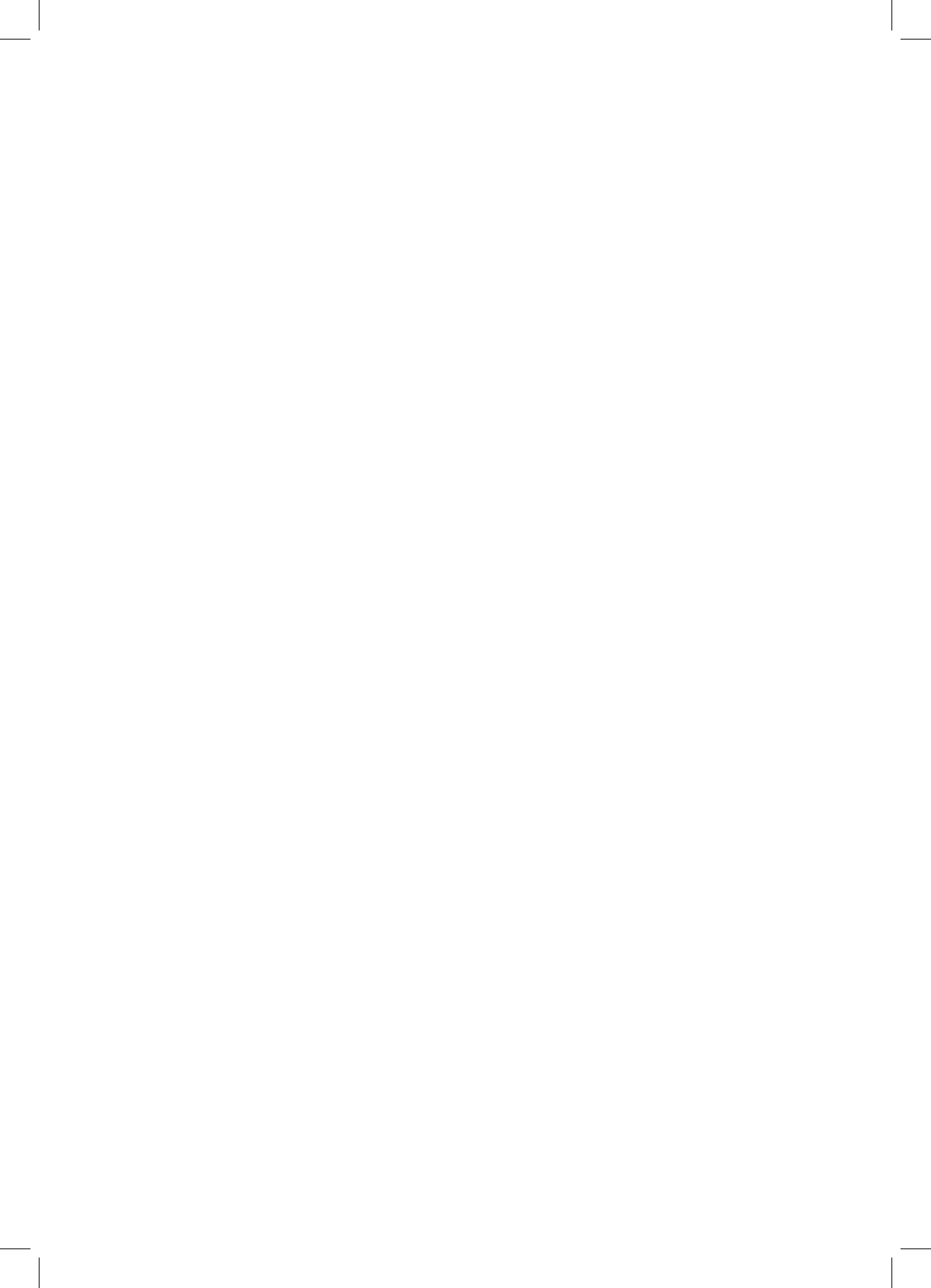
<i>Nathan P. Gibson, David A. Michelson</i> Modeling a Body of Literature in TEI: Bibliotheca Hagiographica Syriaca Electronica	129
<i>Desiree Hebenstreit, Arno Herberth, Laura Tezarek, Christian Zolles</i> Andreas Okopenko: diaries from the poet's estate (hybrid edition)	131
<i>Margaret King, Peter Andorfer, Brian Rosenblum</i> Support for TEI editions in academic libraries. An upcoming study of the Library Publishing Coalition	133
<i>Gaia Lembi, Elli Mylonas, Michael Satlow</i> Bibliography in the Inscriptions of Israel / Palestine Epigraphic Project	135
<i>Sünna Looschen, Ingo Börner</i> Database of Belarusian Periodicals	139
<i>Kiyonori Nagasaki, Tetsuei Tsuda, A. Charles Muller, Masahiro Shimoda</i> Tagging on Buddhist Images via IIIF and TEI encoding	141
<i>Sara Palmer</i> Creating Digital Critical Editions with Readux	145
<i>Ariane Pinche</i> Hyperdonat, digital edition project	147
<i>Claudia Resch, Daniel Schopper, Vanessa Hanneschläger,</i> <i>Eva Wohlfarter, Anna Mader, Nora Fischer</i> Wienerisches Diarium Digital: Unlocking a historic newspaper for interdisciplinary studies with the TEI Guidelines	149
<i>François Vignale</i> EuRED: XML-TEI in database structure	151
<i>Joseph Wang</i> From DOCX via TEI to Literature Map	153
<i>Joseph C. Wicentowski, Magdalena Turska</i> Bringing TEI PM to Foggy Bottom	155
<i>Mary Erica Zimmer</i> TEI in the Archives:	

Unfolding the Bookshops in Paul's Cross Churchyard	157
Demonstrations	
<i>Hugh Cayless, Raffaele Viglianti</i> CETELcean, a JavaScript library for isomorphic TEI to HTML transformation	163
<i>André De Tienne</i> STEP Transcriptor: a comprehensive TEI-XML compliant transcription app	165
<i>Eugene W. Lyman II</i> The Elwood Markup & Text Analyzer (TEMTA)	167
<i>Magdalena Turska, Wolfgang Meier</i> Early Print: annotation module for TEI	169
Workshops	
<i>Piotr Bański, Karlheinz Mörth, Laurent Romary, Andreas Witt</i> Lexical information in the TEI: modeling and markup	175
<i>Stefan E. Funk, Thomas Kollatz, Oliver Schmid, Sibylle Söring, Ubbo Veenster</i> Around the world in a TextGrid day. Edit, manage, publish, and explore your TEI data with TextGrid and DARIAH-DE	177
<i>Martina Scholger, Peter Stadler</i> TEI2German Translatathon	183
<i>Magdalena Turska, Lars Windauer</i> eXistdb: more than a database	185
<i>Tanja Wissik</i> ACDH Tool Gallery 2.2: TEI Basics – A gentle introduction to Text Encoding	187
Special Interest Group meetings	189
Index	191



Introduction





Dear Colleagues,

It is with great pleasure that we present the abstracts of all the presentations of the 16th annual TEI Conference and Members' Meeting to be held in Vienna 2016. This event is intended to cover a wide range of interests but will be an experiment in a number of ways:

TEI 2016 – A conference of openness and diversity

This year's call for papers was formulated by the program committee without a specific thematic strand, encouraging participants to submit proposals reflecting their own particular interests and promising to give equal attention to all incoming papers. The selection of contributions for papers, panel sessions, posters, and demonstrations is therefore a reflection of the wealth and diversity of topics the scientific community is currently engaged in. As a result of this open skies policy, our annual meeting is expected to be a forum for a wide range of disciplines and fields of research. We grouped the manifold contributions into three loose topical strands: The first day is dedicated to "Tools & Techniques", the second day will discuss "Methods & Means", and the third and final conference day will examine "Editions & Editing" in TEI.

TEI 2016 – Copyrights and open licensing

Using ConfTool to prepare the conference proved invaluable in handling the submissions, peer reviews and other tasks, but also ignited a lively debate over the issue of copyright. As a result, the second innovation of TEI2016 is that (following the local committee's suggestion) for the first time all contributions – in digital and print format – will be published under a Creative Commons Attribution license (CC-BY 4.0). This will allow both authors and readers (as stated) to redistribute and adapt these contributions as long as the creators are credited.

TEI 2016 – Evaluating new formats

Another innovation of this years' conference is a series of demonstrations, in addition to the usual panel sessions, workshops, SIG meetings, business meeting, and formal meetings of the Board and the TEI Technical Council. We would particularly wish to draw your attention to the round table discussion "Whither TEI: The Next Thirty Years." This will spark the discussion on the core themes and debates that have characterized the first 30 years of the TEI and help to put in place the plans for the future of the TEI, both in the medium and the long term.

Before finishing up, we would like to take the opportunity to thank all those who were involved in preparing this conference: the program committee members for their engagement, participation, and essential advice in selecting the papers as well as the local committee members for their high professionalism and great commitment over the past months. We also wish to express our appreciation to all participants and presenters for all of the papers, posters, demonstrations, and workshops with which they will become the true protagonists of TEI 2016. On a final note, we would like to formally express our gratitude to our partners for their financial support, namely the City of Vienna, Syncro Soft (<oXygen/> XML editor), eXist Solutions, and appsoft Technologies (Xeditor). We are equally grateful for the commitment of the Austrian Academy of Sciences for supporting the Austrian Centre for Digital Humanities in hosting this event. Welcome to TEI 2016 in Vienna – enjoy the conference!

Martin Mueller
Chair of the program committee

Claudia Resch
Chair of the local committee

Vienna, September 2016

Keynote





Tara Lee Andrews

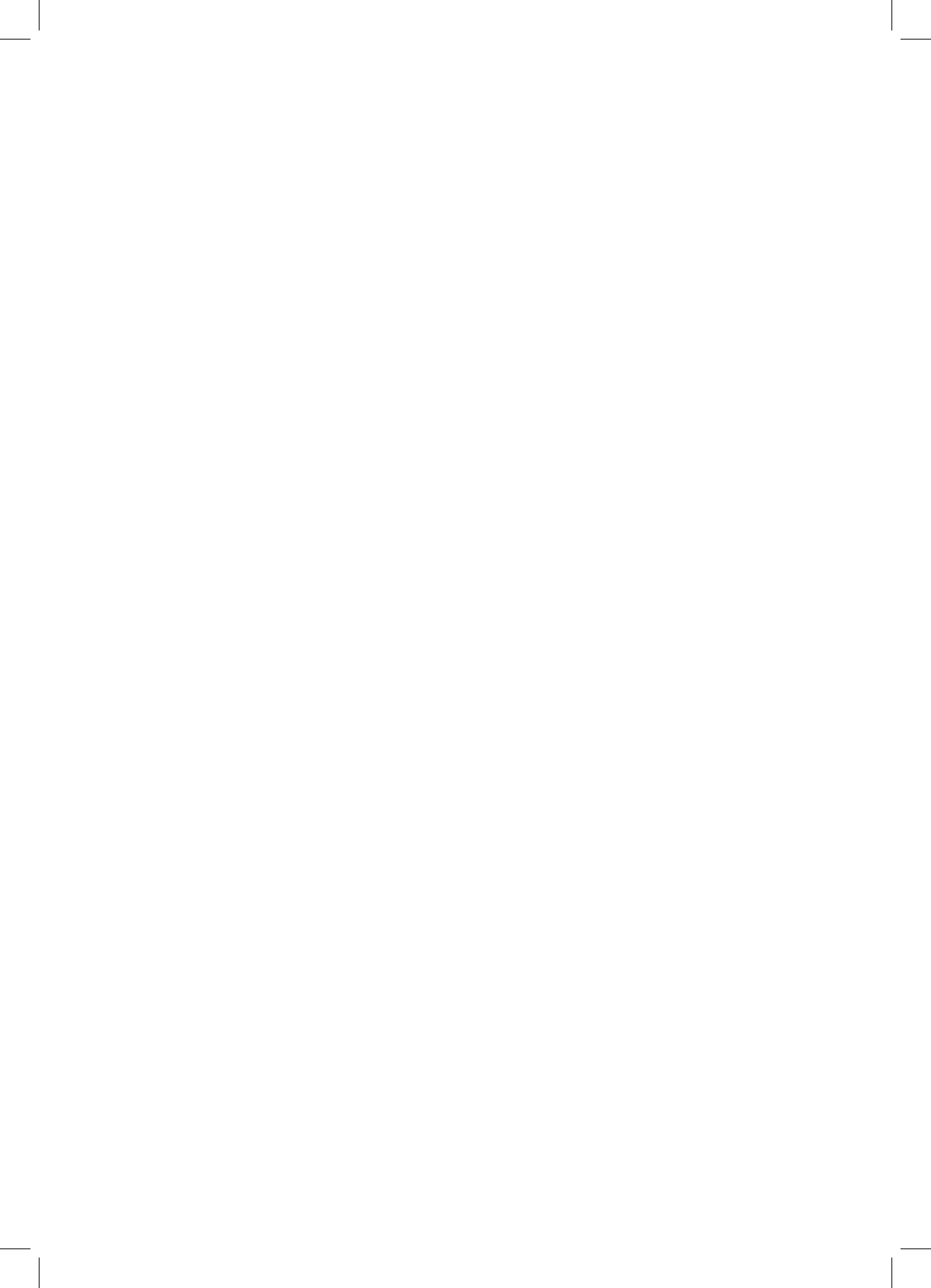
Freeing our Texts from their (Digital Tool) chains

For as long as the TEI has existed, the needs of scholars to interpret and express the texts they encode has sat uncomfortably with the requirements of developers that the code they write be interoperable and usable beyond a single text or a single project. Fundamentally, software development for textual scholarship needs is just as much an act of interpretation as the encoding itself, and interoperability is just as problematic as any exact repetition of scholarly methodology ever is. This is far from being bad news, however! When code is interpretation, it is also scholarship. By breaking away from the concept of code-as-toolbox in favor of code-as-scholarship, we can make a true and decisive argument for the place of the digital in the humanities.



Plenary





TEI Technical Council

Whither TEI: The Next Thirty Years

The triennial “Antonio Zampolli Prize for a singular project or accomplishment” will be awarded to the TEI community at the 2017 meeting of the ADHO in Montreal. This is a wonderful 30th birthday present to the TEI, which was born at a conference at Vassar College in 1987. The award recognizes the work of the many hands that made the TEI a central player in the digital remediation of cultural heritage objects that came to life in print or manuscript form. But complacency is the greatest enemy of institutional health, and the award, however well-deserved, should act as an incentive to think about the next 30 years and the ways in which the TEI can maintain and strengthen its role in an environment that has changed greatly since 1987 and will continue to change in rapid and unforeseen ways.

For this reason, “Whither TEI: The Next Thirty Years” will be the topic of two plenary sessions of this conference. The sessions will be led by members of the TEI Board and Technical Council. Their purpose is to stimulate conversations among the community about intellectual, financial, and organizational aspects of the TEI Consortium and about the best ways of strengthening the TEI in the decades to come.



Papers





Florentina Armaselu

Writing under the Magnifying Glass. Encoding the Text in Progress

Iser¹ draws attention to the dual nature of text, determined by its actual structure and the reader's involvement supplying what is not there. This involvement implies an "active interweaving of anticipation and retrospection" and depends on the boundaries imposed by the written versus unwritten text and on the "virtual dimension of the text", nor the text itself or the reader's imagination, but the "coming together of text and imagination"². In his reflections on photography, Barthes³ defines a similar dynamics, that of the "unspeakable which wants to be spoken", manifested through the "punctum", often a detail which triggers the spectator's subjectivity, resulting in something added to the photograph, an addition of "what is nonetheless already there"⁴.

The present study proposes a view on the coming together of text and imagination, anticipation and retrospection, written and unwritten, unspeakable wanting to be spoken, from a different perspective: that of the text in progress, in its gradual development during the process of writing. The proposal is centered on a textual model (z-text) and inter-

¹ Wolfgang Iser (1974): *The Implied Reader: Patterns of Communication in Prose Fiction from Bunyan to Beckett*. Johns Hopkins University Press, Baltimore.

² *Ibid.* 282, 279.

³ Roland Barthes (1981): *Camera Lucida. Reflections on Photography*. Translated by Richard Howard. Hill and Wang.

⁴ *Ibid.* 19, 55.

face (z-editor)⁵ allowing to keep trace of the evolving text and to explore it “under the magnifying glass” by zoom-in and zoom-out (z-reading). Z-text writing (z-writing) consists of the expansion of already written fragments (z-lexias) by gradually adding new details to them. The model is based on TEI encoding, to each level (phase of development) corresponding an XML-TEI file. The z-lexias are marked-up by `<anchor>` elements and identifiers recording the ancestors-descendants relationships engendered in the process. The paper will discuss the textual model / interface and further ways of analysis / visualization / transformation potentially supported by the encoding, inspired by concepts and studies such as genetic graphs⁶ or informational entropy⁷ applied to linguistic analysis⁸.

⁵ Florentina Armaselu (Vasilescu) (2010): Le livre sous la loupe: Nouvelles formes d'écriture électronique. In Papyrus. University of Montreal Institutional Repository, <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/3964> [2016-07-31]; Florentina Armaselu (2014): The Layered Text. From Textual Zoom, Text Network Analysis and Text Summarisation to a Layered Interpretation of Meaning. In Bornet, Cyril (ed.): Digital Humanities Conference, Lausanne, Switzerland, 7-12 July 2014. Book of Abstracts. EPFL – UNIL. 80-83.

⁶ Lou Burnard; Fotis Jannidis; Elena Pierazzo; Malte Rehbein (2010): An Encoding Model for Genetic Editions. In Workgroup on Genetic Editions. <http://www.tei-c.org/Activities/Council/Working/tcw19.html> [2016-07-31].

⁷ Claude E. Shannon (1948): A mathematical theory of communication. In The Bell System Technical Journal.

⁸ Solomon Marcus (1970): Poetica matematică (Mathematical poetics). Editura Academiei R.S.R., București.

References

- Armaselu (Vasilescu), Florentina (2010): Le livre sous la loupe: Nouvelles formes d'écriture électronique. In Papyrus. University of Montreal Institutional Repository. <https://papyrus.bib.umontreal.ca/xmlui/handle/1866/3964> [2016-07-31].
- Armaselu, Florentina (2014): The Layered Text. From Textual Zoom, Text Network Analysis and Text Summarisation to a Layered Interpretation of Meaning. In Borner, Cyril (ed.): Digital Humanities Conference, Lausanne, Switzerland, 7-12 July 2014. Book of Abstracts. EPFL – UNIL. 80-83.
- Barthes, Roland (1981): Camera Lucida. Reflections on Photography. Translated by Richard Howard. Hill and Wang.
- Burnard, Lou; Jannidis, Fotis; Pierazzo, Elena; Rehbein, Malte (2010): An Encoding Model for Genetic Editions. In Workgroup on Genetic Editions. <http://www.tei-c.org/Activities/Council/Working/tcw19.html> [2016-07-31].
- Iser, Wolfgang (1974): The Implied Reader: Patterns of Communication in Prose Fiction from Bunyan to Beckett. Johns Hopkins University Press, Baltimore.
- Marcus, Solomon (1970): Poetica matematică (Mathematical poetics). Editura Academiei R.S.R., București.
- Shannon, Claude E. (1948): A mathematical theory of communication. In The Bell System Technical Journal.

Piotr Bański, Bertrand Gaiffe, Patrice Lopez, Simon Meoni,
Laurent Romary, Thomas Schmidt, Peter Stadler, Andreas Witt

Wake up, standOff!

The paper will provide an overview of and an update on the ongoing proposal to create a `<standOff>` component within the TEI architecture. It will elicit the conceptual background of having stand-off annotations embedded within a TEI document and the consequences in terms of primary source preservation, multiple annotation views and possible exporting of annotation content into autonomous TEI documents. It will demonstrate the various types of possible use cases ranging from manual annotation to fully automatized information extraction processes and show the importance of implementing, right from the onset, the possibility to use any kind of internal or external vocabulary for representing annotation bodies (e.g. to deal with structural or conceptual annotations). An important prospect here is that the `<standOff>` construct could lead to a simplified development of TEI-aware online services such as Named Entity Recognizers.

We will relate to ongoing initiatives and show the necessity to align with the Web Annotation Data Model (W3C) as well as with the recent introduction of the `<annotationBlock>` element for speech transcription (as part of the work carried out in the ISO standard 24624) as an elementary annotation crystal in the sense of Romary and Wegstein¹. In this context, we will tackle the issue of implicitness in the representation of annotations and open the debate related to the trade-off between having a terse vs. highly flexible model.

¹ Laurent Romary; Werner Wegstein (2012): Consistent Modeling of Heterogeneous Lexical Structures. In *Journal of the Text Encoding Initiative* 3. <http://jtei.revues.org/540> [2016-07-31].

We will end up by illustrating the application that is already made of the current proposal in various projects related to data mining or scientific information, and in particular to the representation of annotated scholarly content.

Further material

Minutes of the January 2014 meeting. <http://download2.polytechnic.edu.na/pub7/sourceforge/1/li/lingsig/Documents/Standoff%20in%20Berlin,%2001.2014/standoff-minutesBerlin2014.pdf> [2016-07-31].

The TEI GitHub ticket. <https://github.com/TEIC/TEI/issues/374> [2016-07-31].

The standOff proposal on GitHub. <https://github.com/laurentromary/stdfSpec> (branch AnnArbor) [2016-07-31].

References

Bański, Piotr (2010): Why TEI standoff annotation doesn't quite work: and why you might want to use it nevertheless. In Proceedings of Balisage: The Markup Conference 2010. Balisage Series on Markup Technologies Vol. 5.

ISO/DIS 24624: Language resource management -- Transcription of spoken language. http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=37338 [2016-07-31].

Pose, Javier; Lopez, Patrice; Romary, Laurent (2014): A Generic Formalism for Encoding Stand-off annotations in TEI. <https://hal.inria.fr/hal-01061548> [2016-07-31].

Romary, Laurent (2015): TEI challenges in an accelerating digital world. In DiXiT Convention week, September 2015. The Hague, Netherlands. <https://hal.inria.fr/hal-01254365> [2016-07-31].

- Romary, Laurent; Wegstein, Werner (2012): Consistent Modeling of Heterogeneous Lexical Structures. In Journal of the Text Encoding Initiative 3. <http://jtei.revues.org/540> [2016-07-31].
- W3C: Web Annotation Data Model. <https://www.w3.org/TR/annotation-model/> [2016-07-31].

Michael Beißwenger, Eric Ehrhardt, Axel Herold, Harald Lungen, Angelika Storrer

Converting and Representing Social Media Corpora into TEI: Schema and best practices from CLARIN-D

The paper presents results from a curation project within CLARIN-D, in which an existing 1MWord corpus of German chat communication¹ has been integrated into the DEREKO² and DWDS³ corpus infrastructures of the CLARIN-D centres at the Institute for the German Language (IDS, Mannheim) and at the Berlin-Brandenburg Academy of Sciences (BBAW, Berlin).⁴ The focus is on the solutions developed for converting and representing the corpus in a TEI format.

The corpus, which has been collected and built in 2002-2008, has originally been annotated using a home-grown XML format that describes the main structural features of chat log files and user postings as well as selected linguistic phenomena of computer-mediated communication (CMC). In order to ensure the sustainability of the resource and its interoperability with the corpus collections already available in CLARIN-D, one important subtask of the project was to define a schema and workflow for remodeling the resource in TEI. Since TEI P5 in its current version doesn't include any models for the representation of CMC

¹ The Dortmund Chat Corpus. <http://www.chatkorpus.tu-dortmund.de> [2016-07-31].

² <http://www1.ids-mannheim.de/direktion/kl/projekte/korpora.html> [2016-07-31].

³ <http://dwds.de/ressourcen/korpora/> [2016-07-31].

⁴ Project description. <http://de.clarin.eu/en/curation-project-1-3-german-philology> [2016-07-31].

and social media genres, the project adopted and extended the modeling suggestions which have been defined and discussed in previous work of the TEI-SIG “computer-mediated communication (CMC)”⁵ and defined a workflow for the automatic, lossless conversion of the source into the target schema.

The target schema⁶ has been tested not only with data from the chat corpus, but also with data from a range of other types of CMC and social media genres (whatsapp interactions, wikipedia talk pages, tweets, use-net posts) in order to provide a useful solution for the encoding of other corpora of that type as well. The schema and conversion workflow will be used for the integration of more CMC and social media corpora into the CLARIN-D infrastructures in the near future.

References

Beißwenger, Michael; Ermakova, Maria; Geyken, Alexander; Lemnitzer, Lothar; Storrer, Angelika (2012): A TEI Schema for the Representation

⁵ Michael Beißwenger; Maria Ermakova; Alexander Geyken; Lothar Lemnitzer; Angelika Storrer (2012): A TEI Schema for the Representation of Computermediated Communication. In Journal of the Text Encoding Initiative 3. <http://jtei.revues.org/476> [2016-07-31]; Thierry Chanier; Celine Poudat; Benoit Sagot; Georges Antoniadis; Ciara Wigham; Linda Hriba; Julien Longhi; Djamé Seddah (2014): The CoMeRe corpus for French: structuring and annotating heterogeneous CMC genres. In Journal of language Technology and Computational Linguistics (JLCL) 29/2. 1-30. http://www.jlcl.org/2014_Heft2/1Chanier-et-al.pdf [2016-07-31]; Eliza Margaretha; Harald Längen (2014): Building Linguistic Corpora from Wikipedia Articles and Discussions. In Journal of language Technology and Computational Linguistics (JLCL) 29/2. 59-82. http://www.jlcl.org/2014_Heft2/3MargarethaLuengen.pdf [2016-07-31].

⁶ The ODD and RNG file for the schema are available in the TEI wiki at <http://wiki.tei-c.org/index.php?title=SIG:CMC:clarindschema> [2016-07-31].

- of Computermediated Communication. In Journal of the Text Encoding Initiative 3. <http://jtei.revues.org/476> [2016-07-31].
- Chanier, Thierry; Poudat, Celine; Sagot, Benoit; Antoniadis, Georges; Wigham, Ciara; Hriba, Linda; Longhi, Julien; Seddah, Djamé (2014): The CoMeRe corpus for French: structuring and annotating heterogeneous CMC genres. In Journal of Language Technology and Computational Linguistics (JLCL) 29/2. 1-30. http://www.jlcl.org/2014_Heft2/1Chanier-et-al.pdf [2016-07-31].
- Margaretha, Eliza; Lungen, Harald (2014): Building Linguistic Corpora from Wikipedia Articles and Discussions. In Journal of Language Technology and Computational Linguistics (JLCL) 29/2. 59-82. http://www.jlcl.org/2014_Heft2/3MargarethaLuengen.pdf [2016-07-31].

Helena Bermudez Sabel

Tomayto, tomahto?

Encoding variant taxonomies in TEI

The inherent flexibility of the digital format has favored the rise of editions that enable access to every witness of a particular textual work. These types of editions might have different goals and seek to answer different research questions, but they usually coincide in drawing attention to the importance of textual variants. To maximize the computational analysis that may be practiced with the variants in different witnesses, a complex taxonomy that reflects the diversity of cases is required.

Many scholars have taken into consideration the recommended TEI method to encode the types of variants – that is, through the attributes `@cause` or `@type` inside the element¹ – and most agree in evaluating it as insufficient.² These attributes are not able to enclose the hierarchy intrinsic to complicated taxonomies or the overlap of classes in an efficient way. However, the TEI Guidelines do offer a valid module for encoding this complex issue: feature structures³. This proposal does not advocate for a controlled vocabulary to categorize types of variants.

¹ TEI Consortium (eds.): 12.1.1 The Apparatus Entry. TEI P5: Guidelines for Electronic Text Encoding and Interchange. 3.0.0. Last updated on 2016-03-29. <http://www.tei-c.org/Vault/P5/3.0.0/doc/tei-p5-doc/en/html/TC.html#TCAPEN> [2016-05-12].

² View discussions in the TEI mailing list searching for key words such as “types of edits” and / or “encoding variants”.

³ TEI Consortium (eds.): 18 Feature Structures. TEI P5: Guidelines for Electronic Text Encoding and Interchange. 3.0.0. Last updated on 2016-03-29. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/FS.html> [2016-05-12].

What it offers instead is a pliable encoding method that allows the editor to include multiple layers of information in each apparatus tagset. As a way to examine the advantages of this method, I will present a practical case in which two “classical” taxonomies (textual: addition, deletion, transposition, and mutation; and substantive against non-substantive types) are combined with a goal-specific multi-layered categorization in a highly efficient way.

Alejandro Bia

TEI document and schema visualization and modeling using mindmaps

In this presentation, we will show the results of the TRACE project in the context of the DHW research group at the Miguel Hernández University. After four years, the TRACE project has served to integrate and evolve several tools created to make working with XML and TEI faster and easier, while favoring mobility and portability by offering them as online services.

This set of tools comprises:

- Tools to graphically visualize and model TEI schemas and document instances.
- Tools for automatic TEI markup from a Markdown-style lightweight markup language.
- Tools to validate, pretty print, improve, and transform XML documents.

In this presentation, we will describe how we use mindmaps to successfully visualize, model, design, modify, import, and export XML-TEI schemas (including DTDs) as well as TEI document instances. Using Freemind, a software tool for drawing mindmaps, and XSLT transformation scripts, we get very manageable, easily comprehensible, folding diagrams from XML sources, which in turn can be edited in a graphical environment and converted back to their original XML format. In this way, we adapted a general purpose mind-mapping tool into a visual tool for XML vocabulary design and simplification. This approach is also very useful for teaching and presentation purposes.

It is frequently said that a good model must be capable of hiding unnecessary detail. The ability to interactively hide / unhide branches of a mindmap diagram, and the automatic allocation of nodes around a center point is what makes Freemind so attractive to our purposes of representing semistructured document structures. User-friendly features for copying, pasting, moving, dragging-and-dropping subtrees make it ideal for visual structure design. For this, we needed a way to import and export several types of schemes. So we implemented transformations for the most popular notations: DTDs, W3C XML Schema, and RelaxNG.

Freemind uses an XML file format, which can be generated using XSLT scripts. We have written several scripts to translate DTDs, XSD, and RelaxNG Schemas to and from Freemind's file format. In the case of DTDs, whose syntax is not actually XML, we used some additional pre / post processing.

These tools will be briefly showcased during the presentation.

Acknowledgements

This work has been developed within the TRACESofTools project: Software Tools for Contrastive Analysis of Texts in Parallel Bilingual Corpora, and has been financed with aid FFI2012-39012-C04-02 from the VI National Plan for Scientific Research, Development and Technological Innovation of the MINECO (Ministry of Economy and Competitiveness of Spain).

References

- Bia, Alejandro; Rodríguez-Sala, Jesús Javier (2016): Building a digital archive for the TRACE project. LIPS 2016: 5th International Library and Information Professionals Summit. New Delhi.
- Bia, Alejandro; Ñeco García, Ramón Pedro; Pérez Beltrán, Jorge (2015): Integración de recursos PLE mediante mapas mentales. In XIII Jornadas de Redes de Investigación en Docencia Universitaria. Alicante, Spain.
- Bia, Alejandro (2015): Down to TEI: use of extended markdown to speed-up the creation of TEI documents. In 15th TEI Conference and Member's Meeting. Lyon, France.
- Bia, Alejandro (2015): Digital Humanities: towards theory and applications. Panel, ESU 2015. Leipzig.

Roman Bleier, Richard Breen

**Documenting Transmission:
The Analysis of the Folk Process using Versioning Machine 5.0**

The Versioning Machine (VM) is a framework and an interface for displaying multiple versions of a text encoded according to the Text Encoding Initiative (TEI) Guidelines. The most recent release of VM contains a new text-audio linking feature, which has originally been developed by members of the Modernist Versions Project and is now a standard component of VM 5.0. The new text-audio feature facilitates parallel reading of a version of a text and at the same time listening to an audio version. Our project attempts to utilize and extend this new feature of VM 5.0 in order to allow the comparison of different versions of a folk song and analyze the “folk process”.

The “folk process” describes the transmission and transformation of literary and artistic material from person to person in both an oral and literary context. As the stories, art, and musical traditions of a community are passed down from generation to generation, they are subject to organic changes in form, context, narrative, and performance. As song forms or tropes travel, they may adopt new musical characteristics, while maintaining a core lyrical or thematic story. In a Digital Humanities context, this raises the question of whether or not a tool like VM 5.0 can be used to document and observe these changes in a way that extends our consideration of these processes.

Our paper will first introduce the new audio feature of VM 5.0. During his MA thesis, Richard Breen encoded several versions of the folk song “The Unfortunate Rake”. Using the VM, he has linked the lyrics to audio

versions of the song, and subsequently was able to observe lyrical comparisons and contrasts. Using this example, the paper will examine the potential of using VM 5.0 to trace the folk process in action.

Roman Bleier, Richard Hadden

**Capturing the crowd-sourcing process:
storing different stages of crowd-sourced transcriptions in TEI**

The Letters of 1916 is a project to create a collection of correspondence from around the time of the Easter Rising, written in Ireland or with an Irish context. The project uses a crowd-sourcing approach to transcription, inviting members of the public to contribute by transcribing letters and correcting those that have already been transcribed. Transcribers use a transcription desk with features borrowed from the Transcribe Bentham project. The backend, based on MediaWiki, stores each saved revision separately, along with relevant metadata.

During our editing workflow, all data is extracted from MediaWiki's database and injected into TEI documents for long-term storage and web presentation. The final crowd-sourced transcription is checked by a member of the editorial team prior to inclusion in our online archive. In addition to storing the final marked-up version of the text, each revision stage is injected and logged in the TEI file. This affords researchers an invaluable resource to study the progress of crowd-encoding, its efficacy, and accuracy over time.

The storage of the different versions of transcriptions in TEI documents is a challenge as, being crowd-sourced, they are seldom well-formed. As an intermediate measure, to enable storage and limited access to the crowd-sourced versions, the `<revisionDesc>` element is employed to record the ID of the transcriber / editor and the time of the revision. The revision itself is "dumped" into the `<revisionDesc>` element inside comment tags to sidestep issues of wellformedness.

This paper will explore more robust solutions for storing and marking-up these XML-like fragments within a TEI document; it will examine possibilities and issues for storing crowd-sourced transcription versions, and how they might be mined for insight into transcription habits.

Jack Thompson Bowers, Thierry Declerck

TEI and lemon: a comparative study on the lexical encoding and interoperability

In this study, we examine questions pertaining to interoperability between the Text Encoding Initiative (TEI) Guidelines for encoding dictionaries¹ and the lexicon model for ontologies (lemon), which has been developed in the context of the W3C OntologyLexica Community Group. The OntoLex Community Group Specification² contains a section dedicated to clarifying the systems relation to other models of lexical encoding, which covers SKOS(-XL), LMF (Lexical Markup Framework), and the Open Annotation standard but lacks a comparison with TEI. We think that such a comparison is needed; also since issues of bringing the TEI into the world of Linked Open Data and making use of ontologies have continued to gain interest within the TEI community.

In our study, we focus on both TEI and the different modules of lemon in aspects of their technical data models, as well as their respective capacities to adequately express certain linguistic features. For this purpose, we have encoded examples in both systems from multiple sources, including historical and dialectal dictionaries, in which we give particular focus on compounds and different types of variation in the lexical entries. In identifying the non-compatible aspects of these systems, we seek to point out and discuss some of the potential linguistic and technical benefits and potential downsides of using each system for one's lexical

¹ <http://www.tei-c.org/release/doc/tei-p5-doc/de/html/DI.html> [2016-07-31].

² http://www.w3.org/community/ontolex/wiki/Final_Model_Specification [2016-07-31].

data, also taking into consideration aspects related to the publication of lexicographic data in the Linked Data cloud.

References

- Declerck, Thierry; Mörth, Karlheinz (2016): Towards a Sense-based Access to Related Online Lexical Resources. In Proceedings of Euralex 2016.
- Schopper, Daniel; Bowers, Jack; Wandl-Vogt, Eveline (2015): dboe@TEI: remodeling a database of dialects into a rich LOD resource. In TEI conference 2015. Lyon, France.

Jack Thompson Bowers, Melanie Seltmann

Exploring data models for heterogenous dialect data: the case of explore.bread.AT!

The project “exploreAT! exploring Austria’s culture through the language glass” aims to explore a collection of heterogeneous 20th century data of Bavarian dialects from the former area of the Austro-Hungarian Empire. Within the project, specific topics such as bread, colors, and plants, and their cultural associations, are investigated in greater detail. Schopper, Bowers and Wandl-Vogt¹ described the process and some unique issues in converting the Database of Bavarian dialects in Austria (DBOE) from a TUSTEP database format to a basic XML format and its supporting resources into an LOD compatible TEI.

This paper discusses issues related to the digital modeling of the data from subproject “explore.bread.AT! exploring Austria’s bread culture dialectally”; among the goals of which are to: extract culturally and linguistically relevant information about bread related topics that may be specific to a given place or time, and enhance the linguistic and semantic description of the dataset using standards, including adding etymological markup and analysis as per Bowers and Romary².

¹ Daniel Schopper; Jack Bowers; Eveline Wandl-Vogt (2015): dboe@TEI: remodelling a database of dialects into a rich LOD resource. In TEI conference 2015. Lyon, France.

² Jack Bowers; Laurent Romary (2016): Deep encoding of etymological information in TEI. Manuscript submitted for publication. <https://hal.inria.fr/hal-01296498> [2016-07-31].

As is the case with much of the rest of the DBOE, this dataset originates from a set of questionnaires, which are a complicated mixture of semasiological (term-based) and onomasiological (concept-based) phrasing, and the content and formatting of the original database entries reflect this. In approaching the modeling of this dataset using markup standards beyond basic XML, we compare and discuss how the data does and doesn't fit within the semasiological model of the TEI dictionary, and where it may fit within a TBXTEI hybrid³ combining certain aspects of the onomasiological model of TBX and the former.

Thus in addition to shedding light on the dialectal data from “explore.bread.AT!”, the issues discussed can be seen as a representative look into core issues present in the remodeling of the larger legacy database.

References

- Bowers, Jack; Romary, Laurent (2016): Deep encoding of etymological information in TEI. Manuscript submitted for publication. <https://hal.inria.fr/hal-01296498> [2016-07-31].
- Romary, Laurent (2014): TBX goes TEI: Implementing a TBX basic extension for the Text Encoding Initiative Guidelines. In Terminology and Knowledge Engineering (TKE). Berlin, Germany.
- Schopper, Daniel; Bowers, Jack; Wendl-Vogt, Eveline (2015): dboe@TEI: remodelling a database of dialects into a rich LOD resource. TEI conference 2015. Lyon, France.

³ Laurent Romary (2014): TBX goes TEI: Implementing a TBX basic extension for the Text Encoding Initiative Guidelines. In Terminology and Knowledge Engineering (TKE). Berlin, Germany.

William Michael Broughton

Encoding Text as a Nexus of Perspective Hierarchies: Case Studies in XML, LMNL, and RDF

It is a fairly easy thing to challenge long standing best practices, especially in a field as increasingly sophisticated as ours is; it is an altogether more difficult thing, however, to actually put that audacious theorizing into practice. I have previously advanced the argument that many of the ills facing the encoding of textual resources are due less to the technical limitations of our practice, but rather to the model of text which underpins it. Instead, I suggest a model of text which extends the (often bemoaned, though still influential) OHCO model, envisioning text not as a single structured hierarchy of content object, but rather as a multi-dimensional matrix of sometimes conflicting, but ultimately interrelated, hierarchies of values. Modeling text – and encoding it – thus would, I argued, allow for nearly seamless data interchange and interoperability, and also for the simultaneous encoding of divergent interpretation within a single interpretative hierarchy – the encoding of dissensus – within a single resource.

However, the time has come to put this theory into practice. What would a text encoded in this way look like? How would it operate? And could any existing encoding scheme or language stand up to the demands I require of it? To answer these questions, I have encoded a single textual source (Doug Zonker's "Chicken Chicken Chicken: Chicken Chicken") using a layered, dimensional approach using three separate encoding languages: TEI-flavored XML, the Layered Markup and Annotation Language (LMNL), and the Resource Description Framework (RDF),

serialized as JSON-LD. While each of the encoding schemes offers unique benefits to the problem, they bring many and varied challenges as well. But both the benefits and the challenges provide valuable insight into the work of reconsidering text as more than the one thing, or even as more than one thing at the same time.

Fabio Ciotti

We are TEI. You will be assimilated!

TEI was originally conceived as a document centric markup language whose domain should have been the digital representation of textual documents in the humanities' domain: mainly literary (at large) sources, archival documents, and language resources¹. At the time when it was developed, no markup-based metadata standards existed, so a great attention was given to the definition of a metadata-oriented section in the encoding scheme, namely the TEI header². The same happened for other kinds of data representation like formulas, graphs, graphic elements, feature structures. In its decennial evolution, even if the data and metadata encoding language ecosystem has changed dramatically, becoming even overpopulated, this approach has been not only confirmed but also reinforced.

During the years TEI has developed markup modules for prosopographical and biographical data³, geographical indexes, taxonomies, and other types of (not strictly textual) structured data; or has "exapted" existing and / or introduced new markup constructs to deal with new representational functions like the expression of Linked Data-like structures, or

¹ C. M. Sperberg-McQueen; Lou Burnard (1995): The design of the TEI encoding scheme. In *Computers and the Humanities* 29/17.

² Lou Burnard; Michael Popham (1999): Putting Our Headers Together: A Report on the TEI Header Meeting 12 September 1997. In *Computers and the Humanities* 33/1-2. 39-47.

³ Eva Wedervang-Jensen; Matthew Driscoll (2006): Report on XML Mark-up of Biographical and Prosopographical data. <http://www.tei-c.org/activities/workgroups/pers/persw02.xml> [2016-07-31].

of stand-off annotations⁴. In general, we can say that TEI has an “assimilation” stance towards any kind of representational need emerging from its users’ community: We could call it a “one markup language does it all” (OMD) approach.

The consequences of this approach are manifold: The TEI encoding scheme has grown gigantic in scope, dimension, and structure, and hence in maintenance complexity; some non-textual modules of TEI, aimed at representing not strictly textual data, are not completely satisfactory for their adopters, and are consequently subject to a strong “pressure” by the user community, resulting in tag abuse or continuous extension and additions requests; the overall consistency of the abstract semantic of the TEI is not assured (a critical analysis of some of these drawbacks in the TEI architecture are in Schmidt⁵).

In this paper, we propose that future development of the TEI should take a more pluralistic approach, permitted by various recent developments of the XML technologies: Namespaces, NVDL (Namespace-

⁴ Christian-Emil Ore; Øyvind Eide (2009): TEI and cultural heritage ontologies: Exchange of information? In *Literary and Linguistic Computing* 29/4. 161-172; Øyvind Eide (2015): Ontologies, Data Modeling, and TEI. In *Journal of the Text Encoding Initiative* 8. <http://jtei.revues.org/1191> [2016-07-31]; Piotr Bański (2010): Why TEI standoff annotation doesn’t quite work: and why you might want to use it nevertheless. In *Proceedings of Balisage: The Markup Conference*. Balisage Series on Markup Technologies Vol. 5; Fabio Ciotti; Francesca Tomasi (forthcoming): Formal ontologies, Linked Data and TEI semantics. In *Journal of the Text Encoding Initiative* 9.

⁵ Desmond Schmidt (2014): Towards an Interoperable Digital Scholarly Edition. In *Journal of the Text Encoding Initiative* 7. <http://jtei.revues.org/979> [2016-07-31].

based Validation Dispatching Language⁶) and multi-schema validations, XInclude.

Recently TEI has extended its capacity to cooperate with external data and metadata representation languages introducing the new element `<xenodata>`, a container element into which metadata in “non-TEI” formats may be placed⁷. We propose some possible extensions of this element, in order to pursue this approach more thoroughly:

1. a special attribute to state the type of metadata vocabulary used (for validation purposes);
2. a flexible method to assert that a set of external metadata (not necessarily in XML format) applies to a specific part of the TEI file: for instance in order to give a MODS description of bibliographic items cited in the text; to express MIX technical metadata of facsimiles images; or to give a formal definition to a `<term>` (or whatever) element by the way of an RDF triple expressing a SKOS based thesaurus;
3. a way to adopt external metadata sets (MODS, VRA, Premis, etc.) to substitute or implement the internal ones.

We advocate, in sum, a wider conception of `<xenodata>` and a more fine grained way to assert the possible relationships between the metadata it contains and their objects. This could be an operational intermediate step towards a more general redesign of the TEI encoding schema.

⁶ ISO (2006): ISO/IEC 19757-4 NVDL (Namespace-based Validation Dispatching Language). [http://standards.iso.org/ittf/PubliclyAvailableStandards/c038615_ISO_IEC_19757-4_2006\(E\).zip](http://standards.iso.org/ittf/PubliclyAvailableStandards/c038615_ISO_IEC_19757-4_2006(E).zip) [2016-07-31].

⁷ TEI Guidelines. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/HD.html#HD9> [2016-07-31].

References

- Bański, Piotr (2010): Why TEI standoff annotation doesn't quite work: and why you might want to use it nevertheless. In *Proceedings of Balisage: The Markup Conference*. Balisage Series on Markup Technologies Vol. 5.
- Burnard, Lou; Popham, Michael (1999): Putting Our Headers Together: A Report on the TEI Header Meeting 12 September 1997. In *Computers and the Humanities* 33/1-2. 39-47.
- Ciotti, Fabio; Tomasi, Francesca (forthcoming): Formal ontologies, Linked Data and TEI semantics. In *Journal of the Text Encoding Initiative* 9.
- Eide, Øyvind (2015): Ontologies, Data Modeling, and TEI. In *Journal of the Text Encoding Initiative* 8. <http://jtei.revues.org/1191> [2016-07-31].
- ISO (2006): ISO/IEC 19757-4 NVDL (Namespace-based Validation Dispatching Language). [http://standards.iso.org/ittf/PubliclyAvailableStandards/c038615_ISO_IEC_19757-4_2006\(E\).zip](http://standards.iso.org/ittf/PubliclyAvailableStandards/c038615_ISO_IEC_19757-4_2006(E).zip) [2016-07-31].
- Ore, Christian-Emil; Eide, Øyvind (2009): TEI and cultural heritage ontologies: Exchange of information? In *Literary and Linguistic Computing* 29/4. 161-172.
- Schmidt, Desmond (2014): Towards an Interoperable Digital Scholarly Edition. In *Journal of the Text Encoding Initiative* 7. <http://jtei.revues.org/979> [2016-07-31].
- Sperberg-McQueen, C. M.; Burnard, Lou (1995): The design of the TEI encoding scheme. In *Computers and the Humanities* 29/17.
- Wedervang-Jensen, Eva; Driscoll, Matthew (2006): Report on XML Markup of Biographical and Prosopographical data. <http://www.tei-c.org/activities/workgroups/pers/persw02.xml> [2016-07-31].

James Cummings

Purifying and Simplifying: Advanced TEI Customization

One of the advantages of the TEI framework is that it is able to be used flexibly and can be modified. It can not only be constrained to be a subset of the existing TEI elements, but also extended to provide markup that the TEI has not (yet) included. The constraining of the TEI scheme happens on multiple levels: the inclusion or exclusion of modules, classes, and particular elements or attributes, but also the values of those attributes and the content models of elements. Unlike most other standards, the TEI framework is also able to be extended, to enable the use of other standards embedded in the TEI or new project specific markup. With the version 3.0.0 release of the TEI Guidelines, there have been two major changes to the underlying TEI customization language, TEI ODD. The first of these are the introduction of “Pure ODD” which answers the long-standing Durand Conundrum by formulating the description of an element’s content not in the RelaxNG schema language but instead purely in TEI ODD. While users may still continue to express their content models in RelaxNG, it is expected that new tools and new customizations will begin to use this so-called “Pure ODD” notation. The second major change has arisen from the TEI Simple project and is the ability to document one or more intended output processing models, in an implementation-agnostic manner, inside your TEI ODD customization. This documentation can then be read by a processing engine to generate the necessary processing workflows. This paper will report on these new aspects of the TEI ODD language while providing a brief refresher for those less familiar with TEI customization.

James Cummings, Jonathan Prag, James Chartrand

Upconversion and Migration: Generating a TEI-EpiDoc Corpus of Sicilian Inscriptions

This paper introduces the TEI P5 / EpiDoc corpus of inscriptions on stone for ancient Sicily called I.Sicily. The corpus aims to include all texts inscribed on stone, in any language, between approximately the seventh century BC and the seventh century AD in Sicily. The corpus currently contains records for over 2,500 texts, and when complete is likely to contain c. 4,000. The corpus is built upon a conversion from a legacy dataset maintained in MS Access and Excel into EpiDoc TEI XML. The XML records are held in an eXist database for xQuery access, and generate other outputs such as a fulltext search using SOLR / Lucene. The corpus and related information (museum list, bibliography) are published as Linked Data, and are manipulated through a RESTful API. The records are queried and viewed through a web interface built with AngularJS and jQuery JavaScript components. Mapping is provided in the browser by the Google Maps API, and ZPR (Zoom, Pan, Rotate) imageviewing is provided by the IIIP image server.

This paper will report on the main conversion of MS Access and MS Excel files into TEI-EpiDoc XML. This conversion uses a combination of existing TEI stylesheets, and customized transformations to generate thousands of individual TEI-EpiDoc files. These incorporate a variety of references to additional references to standardized vocabularies taken from MS Excel files listing canonical entries. This means that individual inscriptions link through to information about places using Pleiades, lists of museums, and epigraphic types, materials, and supports using

the URIs for EAGLE vocabularies. This paper not only reports on the conversion, providing helpful advice on how to undertake such conversions, but also on the project itself.

Michelle Dalmau, Meridith Beck Mink, James Voelkel, John A. Walsh, Wallace Hooper, William R. Newman, Archie Fields III

The Authoritative Bibliography for the Chymistry of Isaac Newton

The Chymistry of Isaac Newton¹ project team has digitized and encoded, following the TEI Guidelines, the complete corpus of Newton's alchemical manuscripts, which total more than 2,000 pages and over one million words. Newton cited more than 5,000 published and unpublished works in these manuscripts; many of his annotations reference items in his own library, as he was an exceptionally dedicated reader of alchemical texts. Newton's extensive citations and annotations provide a window into his alchemical research and practices, and serve as the basis for our authoritative bibliography of his alchemical sources.

The bibliography, also encoded according to the TEI Guidelines, is being developed as both a standalone reference work as well as an integrated resource with the alchemical manuscripts, providing additional context for Newton's citations and florilegia. Once finished, the bibliography will provide complete, structured citations – which often would appear very abbreviated or incomplete in the manuscripts – that can be formatted to comply with modern bibliographic conventions and bibliographic management systems. Our bibliography will also link to digitized online versions of the source texts available through Early English Books Online, HathiTrust Digital Library, and other digital repositories. The citations include quasi-facsimile title page transcription, a technique used for bibliographic description of rare books, to enable richer forms of citation

¹ <http://chymistry.org> [2016-07-31].

analysis. By analyzing the citations, we will be able to date Newton's manuscripts, cluster manuscripts that cite the same or related sources, and, ultimately, generate network graphs that will reveal connections between the cited authors and texts and how they influence Newton's own ideas and work. This paper will provide an overview of the methods and approaches used to compile this authoritative bibliography, tactics for integrating the bibliography with the manuscript collection, and set the stage for computational analysis of the citations.

Stefan Dumont, Susanne Haaf, Tobias Kraft, Alexander Czmiel,
Christian Thomas, Matthias Boenig

Applying Standard Formats and Tools: “Alexander von Humboldt auf Reisen” as an Example for the Collective Subsequent Use of DTABf and ediarum

About the Project

In the project “Travelling Humboldt – Science on the Move” (here: AvH) of the BBAW¹, a digital edition of those writings of Alexander von Humboldt are provided, which emerged from his journeys to America and Russia. The project is supported by TELOTA², which develops tools for the creation of digital scholarly editions, and DTA³, which provides a platform for large TEI corpora of historical texts.

Subsequent Use in and from AvH

The source material is very heterogeneous: Humboldt rearranged and complemented his journals subsequent to his journeys. Nevertheless, it was possible to base the main part of the annotation guidelines on the existing DTABf-M⁴; only a small amount of project specific addi-

¹ Alexander von Humboldt auf Reisen – Wissenschaft aus der Bewegung. <http://www.bbaw.de/en/research/avh> [2016-07-31].

² The Electronic Life Of The Academy. <http://www.bbaw.de/en/telota> [2016-07-31].

³ Deutsches Textarchiv (German Text Archive). <http://www.deutschestextarchiv.de> [2016-07-31].

⁴ Susanne Haaf; Alexander Geyken; Frank Wiegand (2015): 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 8. <http://jtei.revues.org/1114> [2016-05-15].

tions was necessary. The digital edition is created in ediarum⁵, which has been adapted for the DTABf-M and the project specific annotations. The project's encoding demands are examined for their generalizability. If they may serve other projects as well, the respective adaptations are implemented in the DTABf-M and ediarum.BASIS. This way, standards are directly optimized based on their usage within a project.

The resulting standardbased digital resources may in turn be reused in diverse contexts. For instance, the DTABf-conformant texts may be added to the DTA corpora, may there be combined esp. with the apostilles of Humboldt's Cosmos lectures⁶ and from there integrated in the CLARIN infrastructure. With the web service correspSearch, the edited letters can be linked to other letters already published elsewhere in editions⁷.

Conclusion

The current use case is an example for consequent reuse of existing TEI resources, workflows, and tools within multiple projects. This way, efforts are not concentrated on new developments but rather on the improvement of existing standard tools and formats as well as the handling of project specifics. The creation and usage of interoperable TEI resources is an important preliminary in this context.

⁵ Stefan Dumont; Martin Fechner (2015): Bridging the Gap: Greater Usability for TEI encoding. In Journal of the Text Encoding Initiative 8. <http://tei.revues.org/1242> [2016-05-14].

⁶ <https://www.culture.hu-berlin.de/de/forschung/projekte/hidden-kosmos> [2016-07-31].

⁷ <http://correspSearch.bbaw.de> [2016-07-31].

Yasmin Faghihi, Tobias J. Jocham

Islamic Manuscripts in Cambridge Digital Library & the German-French virtual library project Paleocoran

European libraries have been actively collecting manuscripts from the Islamic world since the 16th Century. Until recently, the only way of discovering these collections was through inconsistently produced printed catalogues. In 2009, Cambridge (UL) and Oxford (Bodly) embarked on a joint project to create an online catalogue of Islamic Manuscripts and a structured approach to Islamic manuscript descriptions using TEI. The focus was on creating a standard practice for TEI description of Islamic manuscripts to promote interoperability of data. The result was a union catalogue of Oxford's and Cambridge's holdings, which now includes twelve further institutions (FIHRIST). Content (TEI files) and methodology (schema, practice) were reused and expanded in Cambridge Digital Library, fostering collaborations with external projects. In this paper, I will talk about the challenges and rewards of using TEI for Islamic Manuscript description and about our work in creating a "TEI-community" within the field.

PALEOCORAN aims at bridging the gap between

1. the general history of the Qur'ān as known through Arabic sources and latest paleographic research on the manuscripts and
2. the actual reception of various aspects of the text as documented in the library of the 'Amr mosque in al-Fuṣṭāṭ.

The fragmentary state of early Qur'ānic manuscripts, scattered between various collections, has prevented many attempts at examining thoroughly all manuscript evidence. The digital reconstruction of the Fuṣṭāṭ

collection includes fragments from all over the world, most however being kept in Gotha, Saint Petersburg, and Paris (approx. 11,000 fol.) and unites them in a virtual online library.

Besides a unified cataloguing of its approx. 360 fragments, an important part of the project is focusing on the development of the Arabic script (paleography, letter shapes, diacritical signs, vowel system) and the process of canonization of the Qur'ān within this collection. Thus, the manuscripts will be approached in a multidisciplinary way, combining philology, paleography, codicology, art history, and physico-chemical analyses (ink analysis and 14C). Whilst some of this data can easily be gathered and analyzed with common data models, the encoding of variant readings applied to a certain Arabic word within a manuscript requires more complicated considerations on the design of the XMLschema. Furthermore, a convenient way for entering and editing this data, i.e. Arabic words with multiple markups on single vowel points, by the researchers is needed and has not yet been applied to a similarly huge collection.

References

- Cambridge Digital Library. Mingana-Lewis Palimpsest. <http://cudl.lib.cam.ac.uk/collections/minganalewis> [2016-07-31].
- Cambridge University Library Special Collections. <https://specialcollections.blog.lib.cam.ac.uk/?p=12005> [2016-07-31].

Elena González-Blanco, Antonio Robles-Gómez, Salvador Ros, Gimena Del Río Riande, Roberto Hernández, Miguel Urizar, Clara I. Martínez, Rafael Pastor, Jesús Cano, Agustín C. Caminero

TEIScribe: A graphic tool for composing and testing TEI documents in the context of the EVI-LINHD environment

Digital Humanities can be seen as a boundary discipline that requires cooperation and common agreements and views among many scientific communities.¹ This is the case of Virtual Research Environments (VREs), as they facilitate researchers and users from different communities a place to develop, store, share, and preserve their work.² The first Digital Humanities Center in Spain, LINHD, the Digital Innovation Lab at UNED has started developing EVILINHD, the first VRE for Spanish-speakers.³ The environment offers researchers a collaborative space in the cloud to manage all phases of their projects: the edition process, storage into the database, and text visualization in several output formats, such as HTML. In order to facilitate digital scholarly editing, a specific tagging tool has

¹ Gimena Del Río Riande (2016): Humanidades Digitales. Construcciones locales en contextos globales. SEDICIBlog. <http://sedici.unlp.edu.ar/blog/2016/03/22/humanidades-digitales-construccioneslocales-en-contextos-globales/> [2016-05-24].

² Annamaria Carusi; Torsten Reimer (2010): Virtual Research Environment Collaborative Landscape Study. A JISC funded project. Oxford e-Research Centre, University of Oxford and Centre for e-Research, King's College London. <https://www.jisc.ac.uk/rd/projects/virtual-research-environments> [2016-05-24].

³ Elena González-Blanco (2016): Un nuevo camino hacia las humanidades digitales. El laboratorio de innovación en humanidades digitales de la UNED (LINHD). In Signa: Revista de la Asociación Española de Semiótica 25. 79-93. <https://dialnet.unirioja.es/servlet/articulo?codigo=5476762> [2016-05-24]; Del Río Riande 2016.

been developed as a cloud application implemented with Vaadin, an open-source Java framework and integrated in EVILINHD: TEIScribe.⁴

TEIScribe helps the Digital Humanities community to concurrently label texts with TEI without XML knowledge through a graphic and intuitive design that aims to break language and technological barriers. The tool, which was conceived and developed for the BIESES project in order to boost collaborative work⁵, is based on some of the existing collaborative cloud editors (such as FontoXML, or CWRC writer), so its learning curve is low. The creation, modification, and eliminating of labels and attributes is done with only a few mouse clicks. Since not all projects have the same labeling needs⁶, each text in TEIScribe is linked to a particular scheme that establishes the TEI file structure (an XML file). All documents employed by the application are stored in a NoSQL database, a documental database, namely eXistdb, which organizes the different XML documents and schemes by project, and their correspondence. In this way, the tool greatly simplifies the user's work, since it can automatically detect and highlight mistakes on labels, which do not meet the requirement of associated schemes.

⁴ TEIScribe: Testing session in EVILINHD. <http://www.evilinehd.com:8080/Text-Editor/?session=prueba> [2016-05-24].

⁵ Nieves Baranda Leturio: BIESES project. Biografía de EScritoras ESpañolas. <http://www.bieses.net/> [2016-05-24].

⁶ Elena Spadini (2015): TEI and web-based editing tools. DiXiT Blog. <http://dixit.hypotheses.org/107> [2016-05-24].

References

- Baranda Leturio, Nieves: BIESES project. Biografía de Escritoras Españolas. <http://www.bieses.net/> [2016-05-24].
- Carusi, Annamaria; Reimer, Torsten (2010): Virtual Research Environment Collaborative Landscape Study. A JISC funded project. Oxford e-Research Centre, University of Oxford and Centre for e-Research, King's College London. <https://www.jisc.ac.uk/rd/projects/virtual-research-environments> [2016-05-24].
- González-Blanco, Elena (2016): Un nuevo camino hacia las humanidades digitales. El laboratorio de innovación en humanidades digitales de la UNED (LINHD). In Signa. Revista de la Asociación Española de Semiótica 25. 79-93. <https://dialnet.unirioja.es/servlet/articulo?codigo=5476762> [2016-05-24].
- Del Río Riande, Gimena (2016): Humanidades Digitales. Construcciones locales en contextos globales. SEDICIBlog. <http://sedici.unlp.edu.ar/blog/2016/03/22/humanidades-digitales-construcciones-locales-en-contextos-globales/> [2016-05-24].
- Del Río Riande, Gimena; González-Blanco García, Elena; Martínez Cantón, Clara; Escribano, J. (2016): EVI-LINHD. A Virtual Research Environment for the Spanish-speaking Community. Digital Humanities Conference (DH 2016).
- Spadini, Elena (2015): TEI and web-based editing tools. DiXiT Blog. <http://dixit.hypotheses.org/107> [2016-05-24].
- TEIScribe: Testing session in EVILINHD. <http://www.evilinehd.com:8080/TextEditor/?session=prueba> [2016-05-24].

Vanessa Hanneschläger

**Common Creativity international.
CC-licensing and other options for TEI-based
digital editions in an international context**

The digital space has not only boosted inter- and transdisciplinary, but also inter- and transnational scholarly work in the field of the humanities and beyond. The development of the TEI Guidelines, which provide a common standard for encoding texts across national and disciplinary borders, is one of the most influential reactions to the new form of scholarly work founded on the possibilities provided by the internet. Similarly, the need for a legal framework that allows international cooperation and collaboration has become urgent.

Creative Commons licenses were designed as a reaction to the emerging creative production online, which transcends national boundaries. Their aim is to ease sharing and reusing creative content across the borders of national copyright laws. This mission has been widely successful as CC-licenses are today the most frequently used licenses for digital content among international creatives as well as scholars working in the field of Digital Humanities.

However, when it comes to licensing, scholars are often uncertain about the right choice as CC are not the only option: For instance, ODC licenses are often applied to databases. For scholarly content, DiPP (Digital Peer Publishing with DPPL-licenses) has become popular especially in the German-speaking area. One of the reasons for the latter is that CC-licenses are specifically designed to match the World Intellectual Property Organization WIPO's aims and have to be ported to become

compatible with national copyright legislations, while DiPP was designed in Germany and corresponds to German copyright.

In this talk, the different options for licensing TEI-based scholarly projects will be explored. Currently available markup in TEI to encode responsibility as well as examples of projects using different licenses will be discussed. The main questions to be answered are: Why do we need free licenses? What license best fits my intentions and content? And why would I have to restrict reuse of my data?

References

- Creative Commons. <https://creativecommons.org/> [2016-07-31].
- DiPP - Digital Peer Publishing. <http://www.dipp.nrw.de/> [2016-07-31].
- Klimpel, Paul (2013): Free knowledge thanks to creative commons licenses. Why a non-commercial clause often won't serve your needs. Wikimedia Deutschland / iRights.info / CC DE. https://www.wikimedia.de/w/images/homepage/1/15/CC-NC_Leitfaden_2013_engl.pdf [2016-07-31].
- Klimpel, Paul; Weitzmann, John H. (2015): Forschen in der digitalen Welt. Juristische Handreichung für die Geisteswissenschaften. DARIAH-DE Working papers Nr. 12. Göttingen: DARIAH-DE. <https://irights.info/wp-content/uploads/2015/08/Forschen-in-der-digitalen-Welt-Juristische-Handreichung-Geisteswissenschaften-dwp-2015-12.pdf> [2016-07-31].
- Open Data Commons: Legal Tools for Open Data. <http://opendatacommons.org/> [2016-07-31].

Matthew Lawrence Holford, Huw Jones, James Cummings

msDesc at Oxford and Cambridge

The TEI `<msDesc>` module was deliberately created as a flexible scheme, which would “accommodate the needs of many different classes of encoders”. The potential price of this flexibility is the development of discrete communities of practice among different users of the module resulting in small but significant differences in encoding which complicate interoperability and reuse of metadata. Many such differences can indeed be identified. Not only has the module been relatively widely adopted by digital catalogues of Western medieval manuscripts (the manuscripts for which the module was originally most obviously designed). It has also been adopted by catalogues of non-Western manuscripts and other text-bearing objects, such as Fihrist for Arabic manuscripts and EpiDoc for ancient inscriptions. But there have been few assessments of the overall strengths and weaknesses of the module, areas in which it typically succeeds or fails.

This paper examines the use of the module using as case study a current project at the Bodleian Library, Oxford and the University Library, Cambridge, which is assessing how various cataloguing projects within both institutions have implemented the `<msDesc>` module. The aim is to simplify the technical infrastructure for storing and processing the TEI, to provide clearer guidance and improved workflow for future cataloguers, to improve interoperability within and between the respective institutions, and potentially to provide technical frameworks and metadata standards for other institutions looking to create `<msDesc>` records.

Both institutions have developed a wide range of TEI manuscript catalogues: In addition to Fihrist, both have contributed to the catalogue of Hebrew Genizah texts, and there are catalogues of Tibetan, Sanskrit, and other manuscripts and documents. There are considerable challenges in developing a TEI framework that is flexible enough to accommodate the needs of these different sources, while strict enough to meet the demands outlined above.

Robert Klugseder, Christian Steiner

Cantus Network – a semantically enriched digital edition of libri ordinarii of the Salzburg metropolitan province

This contribution will present a digital and semantically enriched edition of liturgical ordinals, called libri ordinarii (LOi), of the metropolitan province of Salzburg. Digital scholarly editions are still a rarely used method in musicology and especially liturgics. Yet, the digital availability of the many surviving liturgical musical sources, which form an essential part of Austria's cultural heritage, is of great importance.

The research project's aim is to reconstruct and produce a synoptic study of the emergence and development of the liturgy in the Salzburg metropolitan province, based on the LOi in the region. A primary task was the transformation of the transcribed LOi from docx into TEI. The styles available in Microsoft Word and special characters were used in order to create markup in Word and were thus converted into TEI using the XSL stylesheets provided by the TEI community. The final version of the TEI was accomplished with a customized workflow in the OAIS-compliant Asset Management System GAMS (developed by the Center for Information Modelling at the University of Graz) involving a Java based routine.

The TEI is used as the primary database and thus semantic representations in RDF as well as printed versions of the LOi as critical editions were produced directly from the TEI representation. In order to output a traditional printed critical edition, LaTeX and its package reledmac were used to transform the transcriptions into books. The variants had to be collected in a TEI document using the elements of the Critical Apparatus.

Additional emphasis within the project lies on the use of Semantic Web technologies (OWL, RDF), which will help to reveal interconnections between individual entities and differences between the various liturgical traditions.

References

- Klugseder, Robert (2013): Der mittelalterliche Liber ordinarius der Diözese Passau. Entstehungs- und Wirkungsgeschichte. In Studien zur Musikwissenschaft. Beihefte der Denkmäler der Tonkunst in Österreich 57. 11-43.
- Praßl, Franz Karl (1998): Zur Geschichte und Inhalt des ältesten Salzburger Liber Ordinarius (Codex M II 6 der Universitätsbibliothek Salzburg). In Dobszay, Laszlo (ed.): Cantus Planus Papers Read at the 7th Meeting Sopron Hungary 1995, Budapest. 463-472.
- Rouquette, Maïeul (2016): reledmac. Typeset scholarly editions with LaTeX. <https://xml.web.cern.ch/XML/TLLATEST/texmf-dist/doc/latex/reledmac/reledmac.pdf> [2016-07-31].
- Stigler, Johannes; Steiner, Elisabeth (2014-2016): GAMS and Cirilo Client. Policies, documentation and tutorial. <http://gams.uni-graz.at/doku> [2016-07-31].

Judit Kodolányi

Móricz Digital Edition – Possibilities and connections of a correspondence material

The presentation explores the possibilities and characteristics of the digitalization by TEI XML coding of a substantial sized correspondence, based on the lessons of a currently pursued project. In the framework of the Petőfi Literary Museum of Budapest, since February 2016, we have been processing the correspondence of the 20th century Hungarian prose writer Zsigmond Móricz. We have some 3,000 letters to be processed, from the quarter century extending to 1923, and covering half of his artistic career. The aim is the preparation of a digital edition that includes the criteria for a critical edition for professional audiences, such as annotation and text variants as well as the advantages offered by a digital edition, while also offering an enjoyable reading to the general reader. The research regards the correspondence as a special corpus of texts, which consists of various units possessing peculiar metadata and independent references. These create an organic system of multiple connections among the units and with external data. One of the main objects of the programme is the recovery and integration of these connections, and their presentation and accessibility on a publicly accessible webpage, not simply on the level of metadata but also in their embeddedness in the text. We are also concentrating on connecting the product of our work to the major databases.

It is important that the structure of processing is based on the peculiar nature of the accumulation of literary legacies. The lecture intends to demonstrate the formal framework that we have created in the interest

of uniform integration of new units and handling of the description of correspondence. At the present phase, we are working on the specific methods applicable to the material, suitable for scientific markup.

Patrick McAllister

Does this shoe fit?

Applying the TEI guidelines to Sanskrit philosophical texts

The Text Encoding Initiative's Guidelines and recommendations were, for a large part, developed around the notions of European and American textual scholarship concerned with texts produced in the same cultural context. Naturally, this type of work is what the Guidelines fit best. In dealing with literary productions from other cultural environments, one finds that some of these notions have to be revised: Apparently, clear concepts, for example what a quotation or a title is, can become difficult to explain.

I would like to focus on how some notions have to be changed when working with Sanskrit philosophical texts, a sub-genre of the vast field of Sanskrit literature that the TEI Guidelines have only recently been started to be applied to. In this genre, authors often write extensive commentaries on a previous text, citing a passage, phrase, or only a word, and then explaining and discussing it. These commentaries might themselves be commented upon, and so on, so that we can find several layers of commentaries around a central statement. The encoding and critical editing of these types of texts is not well covered by the TEI Guidelines, and there is a need for discussion. I hope a presentation of the main decisions and considerations made so far in a project that is still work in progress can start such a discussion.

Karlheinz Mörth, Daniel Schopper

Wiki2TEI: Wikipedias as sources for language research

Wikipedia has become a synonym for encyclopedic knowledge to a broad public; Wikipedias are more and more being used for a wide range of applications, among others as source material for research projects. For many languages of the world, the respective Wikipedia is the only freely available digital language resource.

To author Wikipedias, a so-called lightweight markup language, Wiki markup, is used which has a simple syntax, which is supposed to ease editing of web-content to be directly translated into HTML. Unfortunately, Wiki markup has a serious drawback: The lack of consistency in its application, which is mainly due to the fact that it is applied manually without the help of programs checking the digital text's structural integrity (wellformedness) and / or logical consistency (validity). While in the past processing of digital texts often proceeded from plain text, XML technologies have become quite pervasive in many applications. Well-formed XML can be processed in many ways and ensures a higher degree of interoperability.

We have seen many projects aiming to convert Wikipedias into other formats. The probably best-known one is DBpedia, the machine-readable extract of Wikipedias' structured portions, which is a cornerstone of knowledge modeling in the Semantic Web. The goal of our project was to put together a workflow that would allow us to transform the texts contained in Wikipedias into a format that would allow us to use our corpus creation and processing tools: Tokenizer, tagger, indexer, and digital reading environment. As all workflow steps in our environment have

been geared towards TEI, we worked on routines to convert Wikipedias into this more expressive, more reusable format.

Our paper will discuss existing tools to perform this task, our own approach in converting Wiki markup into TEI, and it will give examples of how this data can be used for research.

Anja Morgenstern, Agnes Amminger

Biography as compilation: How to edit Georg Nikolaus Nissen's "Biographie W. A. Mozart's" (1828) in TEI

As an implementation of TEI for textual analysis, the editing project of the early "Biographie W. A. Mozart's" (1828) by Nissen, part of the "Digital Mozart-Edition" (DME) at the Mozart Foundation Salzburg, has recently started. The aim of the project is to reveal the structure of the text by way of markup of the manifold sources, which Nissen used for the biography: These comprise primary sources, such as original letters and documents of the Mozart family, secondary sources, such as contemporary literature about Mozart, and original texts written by the author and later editors.

In our presentation, we discuss the special challenges that arise when creating an edition that focuses on how a text was compiled:

1. The encoding of text passages, which often do not correlate with common text structures (e.g. paragraphs, chapters).
2. The definition of different source types and responsible author / editors.
3. Linking the given source definitions to the text passages they refer to.
4. Integrating an exact bibliography of the sources as well as critical annotations for each of the single text passages.

Additional features of the edition include:

1. The linking of each identified source to digital resources such as the online publications of the DME ("Mozart Letters and Documents

- Online Edition”; the “Neue Mozart-Ausgabe Online”) as well as digital collections of libraries and archives.
- 2. Indices of persons, locations, and works, using standardized data (e.g. GND).
- 3. The creation of a complete early Mozart bibliography up to 1828.
- 4. A search function with free and pre-installed options.

Our project fills a gap in Mozart research, presenting an annotated online edition of this important early Mozart biography which was the first to use original letters of the Mozart family, and, on the other hand, secondary texts on a large scale, though mostly without references.

Matija Ogrin, Tomaž Erjavec

Early Modern Slovenian Manuscripts Between Description, Critical Edition and Lexicon

Even though early modern periods of 17th and esp. 18th century are important for the development of Slovenian literature, the manuscripts from these periods have only been given a sketchy treatment or did not enter the account of scholarly evidence at all. Recent research has proved, however, that many Slovenian Baroque manuscripts entail superior literary, cultural or spiritual values, but have never been published in the medium of printed books, because their authors continued – for several reasons – to rely on the manuscript culture as the main medium of their textual oeuvre.

For this reason, new research initiatives have been undertaken for analysis, transcription, and digital processing of these texts. During the last 15 years (2001-2016), several TEI-conforming projects have been launched, which converged in three main methodological approaches.

1. The fundamental research is encompassed in the analysis of the manuscripts as primary sources, expressed in structured <msDesc> elements and arranged together online as the (beta version) Register of early modern Slovenian manuscripts¹.
2. Digital scholarly editions are an established route of text-critical study, processing, and presentation of selected early Slovenian texts

¹ <http://ezb.ijs.si/nrss/> [2016-07-31].

in the eZISS project², where one of the leading principles is our strict differentiation and inclusion of both diplomatic and critical editions of the text in question, which involves interesting issues of both encoding and on-screen processing.

3. The third approach to the same text is still in progress: We try to find an optimized workflow to generate, for each edition, a lexicon of the words and word forms, extant in its text. Here, again, questions of diplomatic and critical word forms complicate the matter a lot, and we can observe – and aim to address – a range of normalizations, going to contemporary standard Slovene word forms and their lemmatization with part-of-speech tagging, as was demonstrated in the IMP project³. We are experimenting with various machine learning methods for the normalizations, concentrating on Character-based Statistical Machine Translation and Conditional Random Fields. The specific aim of the development of these methods is to facilitate the lexicography work, but could also prove useful in preparing new editions. During the preparation of a new edition, the training dataset for the translator between the diplomatic and critical words and sentences could be bootstrapped as the translation progresses.

² <http://nl.ijs.si/e-zrc/index-en.html> [2016-07-31].

³ <http://nl.ijs.si/imp/index-en.html> [2016-07-31].

Brian L. Pytlik Zillig

Shakespeare's N-grams at 400: Comparing Content Themes With Four Contemporaries

TEI-encoded texts can serve easily as units of content for the purposes of text analysis, and may be up-coded easily with part-of-speech (PoS) data. For the present work, plays by Shakespeare and four of his contemporaries (James Shirley, Thomas Middleton, Christopher Marlowe, and Ben Jonson) were processed with Northwestern University's MorphAdorner tool, developed by Philip Burns. XSLT, which is sometimes misunderstood as merely a display or rendering tool, but is also an excellent text analysis language, was employed to query all PoS-tagged data. Moreover, XSLT easily supports the development of custom statistical functions such as pooled standard deviation and Cohen's *d*, functions helpful for comparing standardized differences between the separate data sets of the various authors.

Experimental routines were then developed to examine speeches in each play for the intersection of two sorts of phenomena:

1. parts of speech located at various locations within a word n-gram sequence of a given length, and
2. content themes as identified by the Linguistic Inquiry and Word Count (LIWC) text analysis framework. LIWC includes many content themes, but for this work, the author examined the following: Love, motion, body, anger, death, murder, and happiness.

Once the XSLT query routines were developed, it became a fairly simple matter to ask questions such as these:

Which author in the corpus wrote more 3-grams in which one of the words was a noun, and on the theme of happiness, or religion, etc.?

How different is the occurrence of the religion theme in Shakespeare's 3-grams compared to Jonson's 3-grams, or Marlowe's with Jonson's?

Which comparisons show the largest and smallest differences?

Texts used here mainly originated in the work of the Text Creation Partnership. Martin Mueller offered additional WordHoard texts for this work.

Claudia Resch, Daniel Schopper

**Encoding crime and punishment:
A digital collection of historic broadsheets**

With this contribution, the authors would like to take the audience on a local time travel: The paper introduces a collection of 180 printed single broadsheets that report on capital punishment in 18th century Vienna between 1700 and 1797. These so-called “death sentences” have hardly ever been the object of investigation so far.

Announcing the lives and crimes of the malefactors, the broadsheets show different degrees of detail. The working group aims at covering personal data as a crucial part of the text genre: Malefactors are not only mentioned with their names and age but also with biographical information such as family background, religious denomination or profession. Most importantly, the leaflets contain descriptions of the offences in question, e.g. theft, incest, infanticide amongst others.

A granular metadata scheme based on the namesdates module will allow for detailed analysis and enable researchers to answer questions like: What was the average age of malefactors? Is there a correlation between the delinquents’ social status and their delicts? Can we observe patterns of male and female criminality?

Already during its preparation, the annotated data provides insight into the exemplary life stories of 18th century criminals in Vienna. By applying this scheme to the whole corpus, the project group hopes to enhance its knowledge on historic executions, thus laying empirical, well-grounded foundations for further research – also with the prospect of facilita-

ting the comparability with data from other European capital cities at that time.

References

- Evans, Richard J. (1996): *Rituals of retribution: capital punishment in Germany 1600-1987*. Oxford [et al.]: Oxford Univ. Press.
- King, Peter (forthcoming): "Hanging not Punishment Enough": The Criminal Corpse, the Criminal Justice System and Aggravated Forms of the Death Penalty in England, 1700-1840. Palgrave MacMillan.
- London Lives 1690 to 1800. Crime, Poverty and Social Policy in the Metropolis. <http://www.londonlives.org/> [2016-07-31].
- Spierenburg, Pieter (2013): *Violence and punishment: civilizing the body through time*. Cambridge [et al.]: Polity Press.
- Ward, Richard (ed.) (2015): *A Global History of Execution and the Criminal Corpse*. Palgrave Historical Studies in the Criminal Corpse and its Afterlife. Palgrave Macmillan.

Chris Rogers

Creating a maintainable cataloging workflow and infrastructure using TEI

For some time, we at Bodleian Digital Library Systems and Services (BDLSS) have struggled to maintain a number of TEI-based manuscript catalogues, based on a custom built indexing and display system for which the code has diverged significantly between implementations.

To resolve this, we have embarked on a project to create a new infrastructure and end-to-end workflow for delivering and maintaining TEI-based catalogues moving forward. The aim is to make life considerably easier for both developers and cataloguers. The tools and documentation development as part of the project will be made open-source for other institutions to use and feed into.

As part of this work, we have engaged in an extensive program of user requirements gathering, and are currently in the process of devising the technical architecture of our new solution.

This presentation would explore the results of our requirements gathering exercise, and the insights gained around preferred workflows for catalogue creators, eagerness to engage directly with code, and ideas for exposing the data in novel ways. As part of the requirements gathering exercise, we also conducted a market review, talking to organizations in the UK and the US about how they currently use TEI. We will discuss these findings, and look forward to the technical solution we will be putting in place.

Daniel Schopper, Kira Kovalenko, Thierry Declerck, Eveline Wandl-Vogt

Encoding a Dictionary of Russian Dialects in TEI and linking to LOD Resources

Within a cooperation project between the Russian and Austrian Academies of Sciences, we are investigating the TEI encoding of the Dictionary of Russian dialects, which contains more than 300,000 entries distributed over 48 volumes. The goal of the study is to increase accessibility, interoperability and reusability of this rich source of dialectal data. Our current proposal for a TEI representation consists in encoding the official Russian word as a TEI `<entry>` element and to use the `<cit>` element for each occurrence of a dialect form (within the `<quote>` element). Within the `<cit>` element, we then also include within the `<usg>` element the geo-location for indicating the region in which the dialect form is used. And finally, we include in the `<cit>` block available bibliographical information (`<bibl>`) – in most cases, from which source the dialect word has been collected.

The meaning of the entry is given in the original dictionary in the form of free text. We are currently working on offering more structure to this part of the original entries, with relevant parts tagged as `<name>`s or, if more flexibility is needed (not only proper nouns), “referring strings” `<rs>`. In the context of the `<name>` element, we include then also conceptual information, for example that an entry is the name of a family of plants (... `<name type="botanicFamily">сложноцветных</name>` ...), which we can then link to the scientific name of this family: `<name type="plant" subtype="scientific" key="taxonid:.." xml:lang="la">Taraxacum officinale Wegg.</name>`. This way, we

can easily link the original entry in the dialectal dictionary to taxonomic datasets that are available in the Linked Open Data cloud, and to other language data included in the Linguistic Linked Open Data.

Daniela Monika Schulz, Nils Geißler, Marcello Perathoner

Building a collaborative editorial workbench for legal texts with complex structures

The “Capitularia”-project is concerned with the (hybrid) edition of decrees by Frankish rulers, an important source for various aspects of early medieval European history. Due to the manner of dissemination, these legal texts are only extant within often quite sundry compilations. The modeling of an overarching structure to depict and reference the single textual units in their various manifestations within the manuscripts hence poses one of the biggest challenges for the TEI encoding.

“Capitularia” is a long term project (2014-2029) funded by the North Rhine-Westphalian Academy of Sciences and Humanities and is being prepared in close collaboration with the “Cologne Center for eHumanities” and the “Monumenta Germaniae Historica” as well as other partners. Since the cooperators are scattered, a central platform for internal communication as well as for the distribution of resources among staff is essential to facilitate successful cooperation. WordPress was selected as CMS, serving not only for the web publication, but also as a collaborative editorial workbench. Features already included (e.g. search facilities, multilingualism) were enhanced, new functionalities added, such as the “Capitularia XSL Processor” plugin that allows for the transformation of XML files to HTML within the CMS. This transformation is the final step of a sophisticated XSLT processing pipeline that has been custom-built. The “Capitularia Page Generator” enables automated generation of pages. In order to facilitate the editorial work with the numerous textual witnesses, collation is backed by variance scores available in the backend.

This functionality is based on CollateX with the algorithms included in the “Capitularia Collation” plugin.

The paper will present straits of encoding capitularies as examples of texts with a complex and varying structure, as well as our approach to deal with these issues. We would also like to discuss further ideas for the future development of a collaborative workbench within the CMS.

Martin Sievers

T³ - Typesetting TEI Using TeX. A Plea for High-quality Typesetting in the Humanities¹

Nowadays, a large and increasing number of editions are “born-digital”, i.e. they are a result of applying IT standards, techniques, and software at (almost) all research stages. XML and the TEI in particular (both organization and standard) paved the way for many academic software tools using XML as an exchange and output format.

However, there are still a significant number of projects, which have to step back into the “old world” when the richly annotated data are supposed to be published as a printed book (Figure 1).

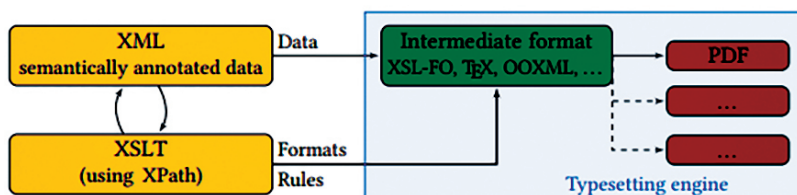


Figure 1: Typesetting XML data for a printed publication. XSLT can be used to transform XML into an appropriate intermediate format. This is then turned into a PDF with the help of a typesetting engine.

Why that? Having a closer look at the latest generation of typesetting engines – both open-source and commercial, it becomes quite obvious,

¹ This submission is dedicated to Sebastian Rahtz, who contributed very important ideas to TeX and the TEI and started to connect both worlds very early.

that the requirements of scientific typesetting have not been taken into account to the needed extent. The main reason is that the “X-technologies”² are primarily industry-driven and therefore corresponding software consequently meets only the requirements of a specific – admittedly large – non-academic target group. Those requirements motivated from a rather academic perspective, however, regularly get much too little attention.

As a consequence, modern virtual research environments like FuD³ or ediarum⁴ started to integrate established non-XML-based open-source typesetting tools like TeX to provide a formatted version for proofreading or even a camera-ready copy for a printed publication. There is even a long-standing direct connection between the TEI format and TeX, namely `teitolatex`, a XSL stylesheet to convert TEI documents into L^AT_EX equivalents. Unfortunately, the XSL stylesheets⁵ don't seem to be very well-known among researchers.

In my talk, I will demonstrate why the “dinosaur” TeX is still needed in a modern XML world and how the community can benefit from more than 30 years of scientific typesetting experience.

² The X-technologies subsume W3C standards, namely XML, XSL and XPath as well as additional related languages and formats like XQuery or XLink depending on the context.

³ FuD – A Virtual Research Environment for the Humanities. <http://fud.uni-trier.de/de/> [2016-07-20].

⁴ ediarum – an easy tool for editing manuscripts with TEI XML. <http://www.bbaw.de/en/telota/software/ediarum> [2016-07-20].

⁵ Cf. <https://github.com/TEIC/Stylesheets> [2016-07-31].

References

- FuD – A Virtual Research Environment for the Humanities. <http://fud.uni-trier.de/de/> [2016-07-20].
- ediarum – an easy tool for editing manuscripts with TEI XML. <http://www.bbaw.de/en/telota/software/ediarum> [2016-07-20].
- Apache(tm) FOP – a print formatter driven by XSL formatting objects (XSL-FO) and an output independent formatter. <http://xmlgraphics.apache.org/fop/> [2016-07-20].
- LATeX – A document preparation system. <http://www.latex-project.org/> [2016-07-20].
- XML-Print: typesetting arbitrary XML documents in high quality. <https://sourceforge.net/projects/xml-printfiles/> [2016-07-20].

Daniil Skorinkin, Evgeny Mozhaev

TEI markup for the 90-volume edition of Leo Tolstoy's complete works

In this paper, we describe our work on an ongoing project titled "Tolstoy Digital". Our chief objective is to convert the 90-volume collected works of Leo Tolstoy into a proper "digital edition" with help of TEI.

With a legacy of more than 46,000 pages of text that collectively contain 14,5 million words, Tolstoy is famed as one of the most productive writers ever. The preparation of the 90-volume print edition started in 1928 (Tolstoy's 100th anniversary) and took three decades, with the last volume published in 1958. Apart from finished works (prose, poetry, drama, essays; schoolbooks), the edition contains numerous drafts, about 8,500 letters, entire volumes of personal diaries, which Tolstoy diligently kept throughout his life, a certain number of facsimile manuscripts and all sorts of editorial comments. A separate volume (No. 91) is entirely dedicated to alphabetic and chronological indexes.

The volumes have been digitized a few years ago, but so far contain little electronic markup. The size and diversity of the edition, along with inevitable inconsistencies in editorial practices, present all sorts of challenges to markup attempts.

One of such challenges are footnotes, of which there are more than 30,000. Among them Tolstoy's own comments, explanations and translations, comments and translations by editors, plus all sorts of technical notes. The latter usually represent various editorial "secondary evidence", e.g. "here Tolstoy wrote word 'A' first, but then replaced it with an unclear

word which is probably word 'B'" or "this phrase was crossed out with a dry pen, most likely by Tolstoy's wife" or "original page contained this addition on the margin".

Obviously, all these footnotes should somehow be distinguished from each other in the TEI markup. As the sheer size of the material suggests some automation is inevitable, currently our efforts are focused on automatic (or at least machine-aided) classification of notes and their subsequent conversion into TEI tags.

Magdalena Turska, Wolfgang Meier

TEI Processing Model Toolbox

As a result of the “digital turn”, editors find themselves in need of more technical competences than were typically considered as core skills for a textual scholar. There are many workshops available where people can learn the basics of TEI encoding in a matter of days, and many are keen to try this method of work in their research. Publishing of the documents encoded with TEI has nevertheless been a sore spot for the TEI community. Simple questions: “How do I get my corpus online? How do I create a PDF?” lead to a lengthy discussion of XSLT transforms and server setups, which usually cools the enthusiasm.

The TEI Processing Model (TEI PM) aims to deal with the transformation part within the TEI ODD language. Still, transformation is just one step when it comes to publishing. Configuring the whole stack of technologies and creating an application necessary to run the website adds several more layers of complexity.

TEI PM Toolbox is a spin-off project of TEI Simple, carried on by Wolfgang Meier, which offers an out-of-the-box publishing of TEI documents. It is an application, which incorporates TEI PM into eXistdb’s web server and application framework, allowing to customize appearance of TEI texts with custom ODD enhanced with processing model instructions. Newest addition to the Toolbox is an application generator allowing to export a freestanding app covering all core features like browsing, search, and generation of various media formats like HTML, PDF, L^AT_EX, or ePUB.

We believe that TEI PM Toolbox is a very practical framework which substantially empowers the editors, especially individual researchers or pilot projects, who struggle to receive sufficient technical support. Following the principle of ODD (one document does it all), it hides many implementation details, still leaving the editors full control over their data and transformation process.

Toshinori Uetani, Sandrine Breuil, Mathieu Duboc, Guillaume Porte

The BVH in Tours and the “Bibliothèques françaises” project

The Bibliothèques Virtuelles Humanistes (BVH, or Virtual Humanistic Libraries¹) is a project run since 2002 by a research team founded by Marie-Luce Demonet in the Centre d'Études Supérieures de la Renaissance (CESR, or Center for Advanced Renaissance Studies: The University of Tours and the CNRS, France).

Its goal is to develop a digital library delivering two types of reliable digital representations, facsimile and text, closely linked together. It offers more than 1,000 facsimiles of typical documents, imprints and manuscripts, of the 15th-17th centuries and about 150 digital editions of French Renaissance texts online. Texts are directly transcribed from the original document and encoded with the XML-TEI standard, adapted to French Renaissance spelling. Our main concern is to merge two workflows of digitization and textual edition following usually two distinct processes, using a single XML schema and to ensure a close correspondence between these two datasets (image-text) incorporating rich metadata.

This long-term program of early modern document corpora is kept partly by several public funding on projects, that is why we diversify projects like the Rabelais website² and the “Montaigne at work” project³, providing accurate digital editions and proposing particular modes of reading and navigation of masterpieces of Renaissance literature.

¹ <http://www.bvh.univ-tours.fr/> [2016-07-31].

² <http://renom.univ-tours.fr> [2016-07-31].

³ <http://montaigne.univtours.fr> [2016-07-31].

Since 2015, the “Bibliothèques Françaises” project combines a digital edition and databases. The texts of the first bio-bibliographical dictionaries of French authors, *Bibliothèque de La Croix du Maine* (Paris, 1584) and that of *Du Verdier* (Lyon, 1585), are encoded in XML-TEI and the bio-bibliographical recordings extracted from the structured text constitute specific databases. These datasets aligned with online databases (biographical data with VIAF; bibliographical records with ISTC, USTC, VD 16, Edit 16 or GLN 15-16) form a dynamic web-space of interoperability and further links to digitized original documents will be a useful tool for data accuracy.

Georg Vogeler

Encoding Text about Things

There is a discussion in the TEI community about the possibilities to encode ontologies with the TEI vocabulary and there is even a SIG on the subject. With the move to P5, the TEI already added several concepts dedicated to the description of “things” (*msDesc*, *person*, *place*, *event* etc.) and thus left the ground of text encoding in a narrow sense. In the last decade, another technological development dealing with digital representations of “things” has taken momentum: the semantic web. The paper will contribute to the reflection about the relationship between the Semantic Technologies proposed by the W3C and the TEI. Its argument starts sticking to the narrow sense of “text” encoding, i.e. the concept of text as a structure of linguistic symbols. The relationship between “things” and text can then be considered as a reference. The TEI offers many methods to establish this kind of references, e.g. the *att.canonical* attribute class. This gives the possibility to express references to “things” which are named individuals. References to classes or analytic concepts can be expressed with the *@ana*, *@type* and the *@inst* attributes, which are globally available.

The power of the W3C Semantic Web standards lies in the highly flexible basic definitions, which still can effectively be processed: Abstract IRI for the symbolization of concepts and “things” and the expression of assertion about the “real world” in simple statements, formalized as “subject-predicate-object” triples. With RDFa, the W3C has proposed a method to integrate formal semantic statements into HTML text.

Projects in which textual information in a digital edition is related to abstract data have shown that easy extraction of highly structured data from texts is a benefit. Theoretically, this approach can draw on the basic insights by Manfred Thaller on source oriented data processing since the 1980s and information theory by Börje Langefors (1995).

The question the TEI community should discuss, is therefore how TEI markup can facilitate the conversion of encoded text into semantic web statements. Is the bunch of @key, @ref, @ana, @inst, link / linkGroup, arc mechanisms more effective than RDFa markup?

At the current stage, the question is open. In my paper, I will try to develop some features of RDFa, which seem to be useful to be considered in the further development of the TEI standards on handling the relationship between text and things.

References

- Langefors, Börje; Dahlbom, Bo (1995): *Essays on Infology. Summing up and planning for the future.* Stockholm: Lund.
- Thaller, Manfred (1988): Data Bases vs. Critical Editions. In *HSR / Historische Sozialforschung* 13/3. 129-139.
- Thaller, Manfred (1989): Datenbasen als Editionsformen. In Schwob, Anton; Kranich-Hofbauer, Karin (eds.): *Historische Edition und Computer.* Graz. 215-241.
- Thaller, Manfred (1988): Gibt es eine fachspezifische Datenverarbeitung in den historischen Wissenschaften? Quellenbanktechniken in der Geschichtswissenschaft. In Kaufhold, Karl Heinrich; Schneider, Jürgen (eds.): *Geschichtswissenschaft und elektronische Datenverarbeitung. Beiträge zur Wirtschafts- und Sozialgeschichte* 36. Wiesbaden. 45-83.

Thaller, Manfred (2012): What is a text within the Digital Humanities, or some of them, at least? In dh2012 – Book of Abstracts, Hamburg. <http://www.dh2012.unihamburg.de/conference/programme/abstracts/beyond-embedded-markup> [2016-07-31].

Dennis Zielke, Alexander Petrus

Automating the validation of TEI metadata processing

This proposal focuses on a TEI metadata ingest procedure, its automated validation and presents the technical implementation as a feature for the data ingest process into the LAUDATIO repository.

LAUDATIO is an open access research data repository for the persistent storage of historical texts and its annotations based on the software framework Fedora. Requirements concerning the metadata structure of historical corpora are identified along with the data depositors. Hence, we will concentrate on the state of the ingest process when depositor and repository manager have already agreed upon the use of a standardized metadata schema. In LAUDATIO, we use TEI P5 as an established and widely used disciplinary metadata format.

The initial validation of (depositor) submitted metadata is a necessary step to ensure a minimum degree of consistency and represents the starting point for any later processing steps. In most repositories, the data-ingest process is form-based. We present a case where depositors have metadata structured in TEI format for which we provide a Graphical User Interface to upload and automatically validate the submitted TEI metadata against standardized schemes. We use the libXML2 library to provide its validation support against the schema. Our proposed module replaces a form-based approach for creating structured repository metadata. The validated metadata is not flawless though.

It relocates and strengthens the necessary cooperation of the repository manager and the depositor. The depositor gets direct feedback and can correct the metadata immediately and independently.



Posters





Peter Andorfer, Dario Kampkaspar

HowTo create your own digital edition web app – A blog

The poster's purpose is the presentation of the blog "HowTo create your own digital edition web app"¹ evolving around a step-by-step tutorial on how to create a "digital edition web application".

Thanks to many summer schools², workshops³, and online-tutorials⁴, the creation of a TEI encoded representation of a text has turned into a manageable challenge for scholars from the humanities. What is still missing are low-entry tutorials on practical approaches to and the technical background of digital editions on the web. The blog launched by this poster tries to be a first attempt to bridge this gap.

The application built and documented by the blog will fulfill the following requirements:

- Customizable (filtering / ordering) **table of contents** derived from the TEI-header information;
- **Transformation of the XML document via XSLT into** (partially customizable) **HTML representations**;

¹ For the time being, the blog is going to be hosted by <http://www.digital-archiv.at> [2016-07-31], though a more stable institutional binding (ACDH-OeAW or HAB) is intended.

² <https://informationsmodellierung.uni-graz.at/de/aktuelles/summer-school-digitale-edition-2016/> [2016-07-31].

³ http://www.tei-c.org/Support/Learn/tutorials.xml#body.1_div.3 [2016-07-31].

⁴ <https://de.dariah.eu/tei-tutorial> [2016-07-31].

- Downloads of **LATeX / PDF** representations generated on the fly;
- **Full text search** and index**based** search;
- **Statistics and (geo)visualizations**;
- **API** to expose the web app's data.

To give a better understanding of the technical backgrounds of general web application techniques and terms, the blog posts will also contain general reflections and discussions of (best) practices in web development and digital edition.

The blog addresses interested scholars and editors without extensive technical background. Likewise, the blog is also interested in opinions and feedback from advanced readers who will be highly encouraged to share their knowledge and expertise by writing posts for the blog.

Therefore, the authors hope that this blog will provide a low entry, open and frequently used platform to read, discuss and engage around the topics digital editions and digital edition web applications.

André De Tienne

STEP Transcriptor and STEP Text Comparator

STEP Transcriptor and STEP Text Comparator are two of several tools being developed to accompany the online Scholarly Text Editing Platform, a general solution that provides a complete transcription, editing, annotation, and layout workflow for scholarly editing projects. STEP Transcriptor has been designed and engineered to help transcribers encode complex texts easily while complying with TEI's manuscript transcription encoding principles. The app (for Mac and Windows) takes care of nearly all of the tedious aspects of TEI encoding. It provides five distinct encoding methods (via pull-down menus, pop-up menus, function keys, keyboard shortcuts, and specialized tagging cursors) that help insert tags, attributes, and values accurately while minimizing typing. Several specialized widgets automatically take care of especially complex encoding situations (including nested alterations and transpositions). STEP Transcriptor automatically creates lists of well-described alterations. It also provides a separate field to enter all sorts of descriptive tags, separately from transcriptive tags. Everything about the app is customizable: the content of menus, the selection of TEI tags, attributes, and values, even tag colors.

STEP Text Comparator is a less ambitious but equally handy tool. It allows editors to compare two versions of a same text at a time, whether variant texts of a same author, or a perfected transcription with its scholarly-edited variant, or a text with and without alterations. The comparison is done paragraph by paragraph, sentence by sentence, sentence segment by segment, and word by word. The software outputs a complete report about all variations (sorted by categories) and automatically creates edit-

able apparatus lists usable for distinct purposes (lists of emendations, of rejected substantives, of alterations) in normal or TEI-XML form.

The poster presentation will display the key functionalities of each software.

Thierry Declerck, Karlheinz Mörrh

Towards a Repository of Senses for Use in TEI encoded Dictionaries

The presentation is based on the observation that information about senses is often repeated in and across larger TEI encoded dictionaries. This has lead us to the idea to set up a repository of senses that can be shared by entries in distinct dictionaries; similar to the ISOcat repository for data categories that can be accessed for encoding part-of-speech and morphological information of lexical entries.

The TEI approach to the encoding of senses is described in the dictionary module of the TEI Guidelines¹. There, an entry is defined as a component-level element (tagged as <entry>) that “contains a single structured entry in any kind of lexical resource, such as a dictionary or lexicon”². A sense (<sense>) is supposed to group “together all information related to one word sense in a dictionary entry, for example definitions, examples, and translation equivalents”³. As such a sense is a component of an entry or of elements of an entry, like homonyms.

There are no defined restrictions as to how to codify the content of the sense, and all possible string characters seem to be allowed. This fact renders the comparison of senses across lexicons difficult, if not impos-

¹ TEI Consortium (2016): TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 3.0.0. <http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf> [2016-04-30].

² Ibid.,276.

³ Ibid.,278.

sible. In general, we do not want to rely on string matching for stating a relation between senses included in different entries. We advocate the creation of a repository of senses, which can be referred to (and shared) by entries in TEI dictionaries. Our experiments made use of technologies such as SKOS-XL, LMF and lemon. To establish efficient linking mechanisms, we made use of the TEI `<ptr>` element.

References

TEI Consortium (2016): TEI P5: Guidelines for Electronic Text Encoding and Interchange. Version 3.0.0. <http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf> [2016-04-30].

Greta Franzini, Peter Andorfer

The Digital Editions Catalogue App: A web application to browse, curate and analyze digital editions

Since 2012, Greta Franzini's "Catalogue of Digital Editions" has been gathering digital editions in an attempt to survey and identify best practice in the field of digital scholarly editing.¹ Other cataloguing initiatives did not, and do not, provide the granular analysis of features necessary to understand the rationale and methodology behind the creation of an edition.² The Catalogue is useful as it provides an accessible record of standards (including TEI) and building tools used, and thus an insight into past and present projects. Originally published as an open Google Sheet, in 2015 the Catalogue moved to GitHub as a comma-separated-value file in order to take advantage of its stable versioning system. This format lends itself well to data download and contribution but not to efficient browsing. In 2016, a web application was developed to deliver the .csv Catalogue data to users in an interactive and user-friendly manner.³ This Django-based application allows users to browse, search, filter and order the Catalogue around their research interests, but also to curate existing data and create new entries through a custom made web front-end. The application also provides statistics and visualizations

¹ Greta Franzini, S. Mahony, and Melissa Terras (forthcoming): 'A Catalogue of Digital Editions'. In Pierazzo, Elena and Driscoll, Matthew James (eds.): *Digital Scholarly Editing: Theories and Practices*. Open Book Publishers. <http://www.openbookpublishers.com/product/483/digital-scholarly-editing--theories-and-practices> [2016-07-31].

² See https://github.com/gfranzini/digEds_cat/wiki [2016-05-14].

³ See https://github.com/acdh-oeaw/dig_ed_cat [2016-05-14].

derived from the data (e.g. map visualizations), as well as the possibility to download and query it via an API. With the Digital Editions Catalogue Web Application, the authors wish to establish a solid and usable platform to collect homogeneous and structured information about digital scholarly editing. The 2016 TEI Conference would provide the perfect setting for an official launch of this collaboration.

Nathan P. Gibson, David A. Michelson

Modeling a Body of Literature in TEI: Bibliotheca Hagiographica Syriaca Electronica

The Syriaca.org project publishes born-digital reference works in TEI. “Bibliotheca Hagiographica Syriaca Electronica”¹ is a set of metadata records describing a particular genre of Middle Eastern literature: saints’ lives written in the Syriac language. This poster describes

1. the process of transforming a legacy (tabular) dataset into TEI and
2. the decisions made in creating a schema for encoding and modeling metadata about a body of literature. Syriaca.org is now revising this TEI schema for an expanded project which will describe all of Syriac literature, a corpus which has had little previous scholarly classification or description.

Syriaca.org considered a number of methodological questions, which arose from using TEI to model “works” (in the sense defined by FRBR²). While TEI is frequently used to represent the content of specific manuscripts (or other textbearing objects) and create editions of works, we found that the use of TEI for born-digital metadata records was less common. We found a number of advantages in using TEI over other metadata or bibliographic formats (e.g. MARC, Dublin Core). In addition, we made encoding decisions about how to apply the TEI guidelines to the following questions:

¹ <http://syriaca.org/bhse> [2016-05-14].

² http://archive.ifla.org/VII/s13/frbr/frbr_current3.htm#3.2 [2016-05-14].

1. How can TEI model relationships among works, especially different recensions of the same work or whole-to-part relationships?
2. How can TEI express the relationships of works to their manuscript witnesses, editions, and translations?
3. How can TEI express the relationship of works to persons with alleged authorial or editorial responsibility for it, whether genuine, pseudonymous, anonymous, or attributed?

This poster presents our preliminary schema for modeling a body of literature in TEI, for which the authors actively seek feedback, suggestions, and criticism. We are particularly interested in learning from similar projects which have used TEI to map or model metadata about large and heterogeneous literary canons.

Desiree Hebenstreit, Arno Herberth, Laura Tezarek, Christian Zolles

**Andreas Okopenko:
diaries from the poet's estate (hybrid edition)**

In the year 2012, the Literary Archives of the Austrian National Library acquired the rights to the literary estate of the Austrian poet Andreas Okopenko (1930-2010), winner, among others, of the Grand Austrian State Prize for Literature (1998) and the Georg Trakl Prize (2002). His estate is a comprehensive collection of diary notations, typescripts, notes, correspondence, and other material. It has been partly arranged by the poet himself, and its classification has been informally continued by the Literary Archives. Regarding text-genetical (1.), poetological (2.) and literary historical (3.) aspects, the diaries are the core of the 28 boxes of the poet's estate (which has to be matched with the 25 boxes of the pre mortem bequest).

Throughout his life, Andreas Okopenko kept a position outside of all existing literary trends. This means that the estate is important in at least three ways: It allows us to compare his personal writings with the typescripts and text variants and to lay the foundations for a comprehensive text-genetical study and interpretation for Okopenko's experimental work – ranging from literary small forms to essays and radio plays, up to his best known avant-gardist novels *Lexikon-Roman* (1970) and *Kindernazi* (1984) (1.). Furthermore, Okopenko's writing is characterized by providing a feeling of "personal" realism (Concretionism) and by a high level of cultural self-reflection. As such, the estate stands as a formidable individualistic testimony of contemporary Austrian history (2.). Finally, Okopenko turned out to be a reserved but rigorous observer

of the literary scene in Austria's Second Republic. Therefore, the diaries in particular will provide different views on the formation and the local value of Austrian literature (3.). The first-range importance of the literary estate as a whole can be found in the areas of text-genetical and interpretational research and in literary and contemporary history. Okopenko's highly personal perceptions and awareness of his own historical context make him a distinguished "archaeologist", "archivist" and "chronicler" of his time.

Above all, the first-range importance of the literary estate can be found in the area of Digital Humanities. Regarding Okopenko's pioneer work in hypertext-poetry, the project will focus on an annotated digital edition of the diaries (following the TEI Guidelines) connected to the website of the Austrian National Library in a sustainable manner and with unlimited access. In addition, the most significant parts of the annotated diaries will be printed. A close collaboration between the Department for German Studies of the University of Vienna (Univ.-Prof. Dr. Roland Innerhofer) and the Literary Archives (Priv.-Doz. Dr. Bernhard Fetz) will guarantee the optimal valorization of the estate and will represent Okopenko's writing in a contemporary light.

Margaret King, Peter Andorfer, Brian Rosenblum

**Support for TEI editions in academic libraries.
An upcoming study of the Library Publishing Coalition**

This poster will present the first outcomes of an ongoing study that aims to address the extent to which digital scholarly editions (and in particular TEI encoded editions) are supported within member libraries of the Library Publishing Coalition. The sample includes more than 100 academic libraries, primarily in the United States as well as some around the world. This organization aims to promote innovative publishing services in these university libraries. Members therefore form a representative sample to address this question.

Library-based publishing has become increasingly common among academic research libraries, as the Library Publishing Directory demonstrates. However, there are not clearly established organizational and service models for these publishing programs. Furthermore, while many of them are focused on scholarly journals, it is unclear how many of these programs are supporting the production and publication of digital scholarly editions making use of the TEI schema, nor is it clear what kinds of resources are required for them to support such efforts.

The survey first aims to determine how many of these academic libraries understand digital scholarly editions to be those making use of the TEI schema and how many have other conceptions. The next objective is to elicit what kinds of support these libraries offer to digital editors including services, publication, preservation and access. Our final aim is to identify the success factors that motivate or enable libraries to offer support to editors and what the primary obstacles are which hinder support

programs. The poster will discuss the initial results and trends of this survey and will give predictions about the findings and recommendations, addressing key issues affecting the publication, sustainability and access of TEI encoded editions.

Gaia Lembi, Elli Mylonas, Michael Satlow

Bibliography in the Inscriptions of Israel / Palestine Epigraphic Project

The Inscriptions of Israel / Palestine project (IIP) has been aggregating and transcribing previously published inscriptions and translations from Israel / Palestine from the Persian period through the Islamic conquest (ca. 500 BCE - 640 CE). There are about 15,000 inscriptions, written primarily in Hebrew, Aramaic, Greek, and Latin, by Jews, Christians, Greeks, and Romans. The project began under the direction of Michael Satlow at the University of Virginia in 1997, and has been located at Brown since 2002. Inscriptions are encoded using the EpiDoc Guidelines, and have extensive metadata that allows them to be searched and browsed on the web via a faceted interface.

Following practices described in the EpiDoc Guidelines¹, IIP has been maintaining bibliographic entries in a master bibliography, and referring to these from individually encoded inscriptions. Since IIP has always employed student encoders, many of the selection criteria cited by Bański et al.² apply to this project as well: Bibliographic data entry and editing has to be collaborative, web based and unambiguous. As a multilingual project, IIP also records Hebrew and Greek titles in their original character set.

In an earlier stage of the project, the bibliography was handled in a customized SQL database, which provided useful features such as controlled

¹ <http://www.stoa.org/epidoc/gl/latest/supp-bibliography.html> [2016-07-31].

² Piotr Bański; Stefan Majewski; Maik Stührenberg; Antonina Werthmann (2012): Building and Maintaining the TEI LingSIG Bibliography. In Journal of the Text Encoding Initiative 3. <http://jtei.revues.org/486> [2016-07-31].

lists of journal abbreviations. When the project was upgraded in 2013, we decided to replace the custom bibliographic database with Zotero, which did not require maintenance on our part and had a versatile API, which allowed us to access citations directly when displaying an inscription in our interface. IIP assigns ID numbers to bibliographic entries, following the form [IIP0000], and uses the ID to reference a citation. Zotero also assigns ID numbers to bibliographic entries. However, these are automatic, and function as database keys, so they are essentially not part of the citation data: We decided therefore to store the IIP ID in the Zotero “Loc in Archive” field, on the premise that it is our project catalog number. In order for the Zotero API to retrieve a particular citation, we also save it as a Zotero tag for the entry, since the Zotero API can only access tags and a restricted number of bibliographic fields. Unlike the LingSIG bibliography, we didn’t need to convert our bibliography to TEI in order to integrate it into our inscription files. It was enough that the Zotero API, using the Citation Style Language (CSL), an open language used by a variety of citation managers, could provide formatted citations for our display.

Our original custom bibliographic system provided especially useful features for ensuring data integrity such as authority lists for journal names and abbreviations and specialized geographic tagging that made it possible to view all items pertaining to a particular location. It did not handle particularly sophisticated bibliographic structures however because we had customizations for each new type of reference. We lost these features when we moved to Zotero, but gained a different set of affordances such as the ability to handle more complex bibliographic structures, to use and customize Zotero’s formatted output. Zotero, as a native bibliographic system, can handle more bibliographic types than our own database could, although some complex older scholarly publications are still problematic.

We feel that our choice of an external bibliographic system – particularly Zotero, which is freely available, independently searchable, and can act as a re-useable, standalone bibliography of sources for inscriptions from Israel / Palestine – is an efficient choice that allows us to work with different types of material using the tools most appropriate for each. In the future, when the IIP corpus is complete, it will be necessary to incorporate full bibliographic entries as `<bibl>`s into each encoded inscription for archival purposes.

References

- Bański, Piotr; Majewski, Stefan; Stührenberg, Maik; Werthmann, Antonina (2012): Building and Maintaining the TEI LingSIG Bibliography. In Journal of the Text Encoding Initiative 3. <http://tei.revues.org/486> [2016-07-31].
- Citation Style Language. <http://citationstyles.org> [2016-07-31].
- Elliott, Tom; Bodard, Gabriel; Mylonas, Elli; Stoyanova, Simona; Tupman, Charlotte; Vanderbilt, Scott et al. (2007-2014): EpiDoc Guidelines: Ancient documents in TEI XML (Version 8). <http://www.stoa.org/epi-doc/gl/latest/> [2016-07-31].
- Rotger, Carlos (2015): Using the Zotero API to Render Formatted Bibliography on a Webpage. Center for Digital Scholarship News. <http://library.brown.edu/cds/2015/06/29/zotero-api-web/> [2016-05-12].
- Michael Satlow (ed.) (2002-2016): Inscriptions of Israel/Palestine. <http://www.brown.edu/iip> [2016-07-31].
- TEI Consortium (eds.) (2016): 3.11 Bibliographic Citations and References. TEI P5: Guidelines for Electronic Text Encoding and Interchange. 3.0.0. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/CO.html#COBI> [2016-07-31].
- Zotero. <http://www.zotero.org> [2016-07-31].

Sünna Looschen, Ingo Börner

Database of Belarusian Periodicals

The proposed poster presents the prototype of the “Database of Belarusian Periodicals”, developed at the University of Oldenburg in cooperation with the Belarusian State University, Minsk.

The digitization of four central literary periodicals (Maladnyak, Polymya, Uzvishsha and Kalos’sse) from 1920 / 1930ies Belarus aims at two main goals:

1. a broad and simplified accessibility of central literary institutions of that time for the scientific community – an important step towards a systematic scientific evaluation of this matter;
2. a manageable database for detailed and possibly statistical analyses of various aspects concerning the institutional landscape, including, but not limited to, the following: the inter-relation of the periodicals’ authors, inter-dependencies of publication place, ideological framework and literary and / or aesthetical standards, frequency of publications, distribution of literary subgenres, etc.

Data capture is performed in XML on the basis of the TEI Guidelines. The poster specifically discusses the application of the Guidelines to the encoding of table of contents of periodicals, which constitute the projects’ main data source.

The database itself is implemented as an application for the open-source XML database eXistdb. An extensive range of search queries can be performed on the collection, which allow to obtain quantitative characteristics of the periodicals and their contributing authors.

Regarding the project's theoretical background (i.e. addressing philological questions about Belarusian literature – which up until now has received limited international attention – from a field theoretical perspective established by French sociologist Pierre Bourdieu¹), the database promises new findings and insights on processes of value assignment and consecration in a less stable literary field². On a broader level, it seeks to generate objective material for the historiographical evaluation of literary processes.

References

- Bourdieu, Pierre (2014): *The rules of art: genesis and structure of the literary field*. Reprint Cambridge: Polity Press.
- Kohler, Gun-Britt (2015): *Westbelarussische Literatur und Belarussisches Literaturfeld (1921-1939). Versuch einer Standortbestimmung*. In *Wiener Slavistisches Jahrbuch, Neue Folge* 3. 135-176.
- Van Rees, Kees (2012): *Field, Capital and Habitus: A Relational Approach to 'Small' Literatures*. In Kohler, Gun-Britt; Navumenka, Pavel I.; Grüttemeier, Ralf (eds.): *Kleinheit als Spezifik: Beiträge zu einer feldtheoretischen Analyse der belarussischen Literatur im Kontext 'kleiner' slavischer Literaturen*. Oldenburg: BIS-Verlag. 15-56.

¹ Cf. Pierre Bourdieu (2014): *The rules of art: genesis and structure of the literary field*. Repr. Cambridge: Polity Press; Kees Van Rees (2012): *Field, Capital and Habitus: A Relational Approach to 'Small' Literatures*. In Gun-Britt Kohler, Pavel I. Navumenka, Ralf Grüttemeier (eds.): *Kleinheit als Spezifik: Beiträge zu einer feldtheoretischen Analyse der belarussischen Literatur im Kontext 'kleiner' slavischer Literaturen*. Oldenburg: BIS-Verlag. 15-56.

² Cf. Gun-Britt Kohler (2015): *Westbelarussische Literatur und Belarussisches Literaturfeld (1921-1939). Versuch einer Standortbestimmung*. In *Wiener Slavistisches Jahrbuch, Neue Folge* 3. 135-176.

Kiyonori Nagasaki, Tetsuei Tsuda, A. Charles Muller, Masahiro Shimoda

Tagging on Buddhist Images via IIIF and TEI encoding

In this poster presentation, we would like to show an implementation of IIIF and an experiment of TEI encoding for a series of Buddhist images and text in the Taishō Tripitaka, which consists of 12 volumes originally published in 1933.³ The series includes images of Buddhist saints, mandalas, rituals, and so on. The implementation enables searching for the images by keyword in English as well as Japanese. The search results are cropped by the IIIF Image API. From this result window, users can make three kinds of selections:

1. to see large images by clicking each image,
2. to see zoomable images by clicking a page number,
3. to see zoomable images in parallel.

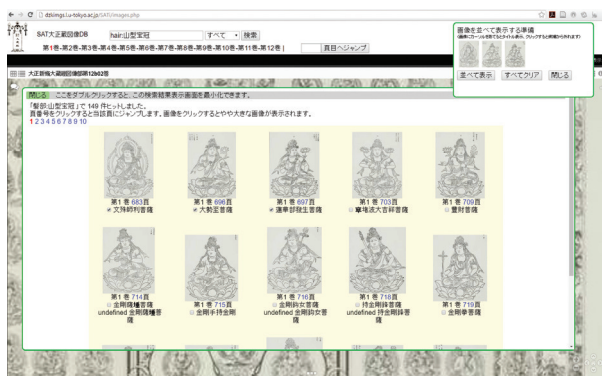


Figure 1

³ <http://dzkings.l.u-tokyo.ac.jp/SATi/images.php> [2016-07-31].

By clicking a checkbox of cropped images, the small thumbnail of the image is listed on a small cart. Users can add more thumbnails onto the cart in the same way (Fig. 1). After that, when a user clicks the “reveal all in parallel” button, the listed images are shown in parallel on Mirador windows (Fig. 2). Moreover, annotations are shown on each tagged region.

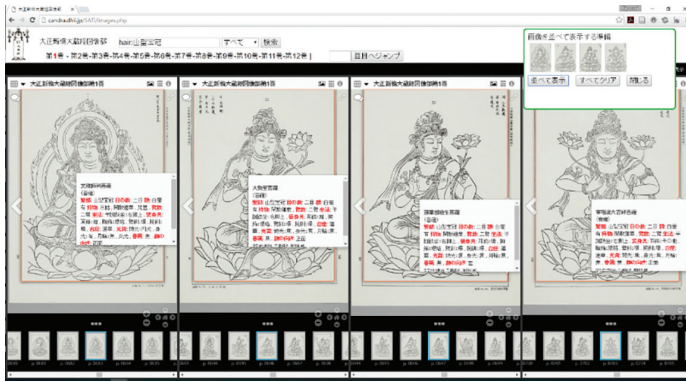


Figure 2

The annotations explain hair style, sitting style, type of chair, belongings, and so on. Then, the user can search from this annotation window by clicking an attribute (Fig. 3).

The user can also see an object in one window like this. The annotations are not yet completed, the number is around 5,200, about 20 percent of all images so far. They were inputted by around 30 researchers of art history to promote research of Buddhist images from the viewpoint of art history through a web collaborative tagging system. Moreover, we are trying to encode the images according to the TEI Guidelines as parts of books. We will present the attempt.



Figure 3

Sara Palmer

Creating Digital Critical Editions with Readux

Readux¹, a reading and annotation platform for Emory University's digitized books, supports exporting an annotated edition as both TEI-encoded data and as a fully featured standalone website. One of the primary objectives of the platform is to support the creation of digital critical editions of texts in which the visual image of the page is essential to scholarly analysis. Using the TEI standards for encoding digital facsimiles, Readux provides an environment in which scholars can add multimedia annotations to selections of text or areas of a page. This environment requires transformation of multiple OCR metadata formats into TEI surfaces and zones as well as the conversion of user annotations made in Markdown to notes anchored to zones in the facsimile. Scholars can choose to export their annotated editions as TEI XML or as a bundled website site built with Jekyll that includes the TEI datafile. Readux also supports publishing directly to GitHub pages allowing users to easily publish an annotated edition to a freely hosted site. Adherence to TEI standards is crucial to Readux's mission to support digital scholarship in a durable archival format that can be adapted to new platforms.

¹ <http://readux.library.emory.edu> [2016-07-31].

Ariane Pinche

Hyperdonat, digital edition project

Hyperdonat is a digital project created to deal with complex textual tradition. Born in Lyon, in the HiSoMA laboratory, from the collaboration of B. Bureau and C. Nicolas, Latin professors, Hyperdonat project was first a digital edition of Donat's commentaries, which have a complex philological history that led editors to dedicate most of their "paper" edition to the apparatus making the Latin text mostly unreadable. The project had as first aims to offer the first complete critical edition and a way to visualize those texts easily with its apparatus through the uses of new technologies.

These last years, the project has taken on a new path and tried to find the basement of an editing method for complex TEI digital editions. So, Hyperdonat project proposes to create from the different witnesses of a textual tradition a scientific edition with multiple views, an interface to create your own virtual witnesses and another to compare different versions of the text. The method is thinking to be applied in a team for a collaborative work where each manuscript is encoded separately to avoid the influence of the other versions. This method allows multiple levels apparatus (semantic, structure, layout, graphic) to take account of the influence of the different layout in manuscript. So this method shows all the philological approach to the reader, but can also be used as a help for editors.

Claudia Resch, Daniel Schopper, Vanessa Hanneschläger, Eva Wohlfarter, Anna Mader, Nora Fischer

Wienerisches Diarium Digital: Unlocking a historic newspaper for interdisciplinary studies with the TEI Guidelines

This poster presents the pilot study for a digital edition of the newspaper “Wienerisches Diarium”, carried out by the Austrian Centre for Digital Humanities in cooperation with the Institute for History of Art and Musicology of the Austrian Academy of Sciences.

The “Wienerisches Diarium” was a periodical published in Vienna from 1703 onwards, with its successor, the “Wiener Zeitung”, still appearing to this day. Founded with support of the Austrian emperor’s court and provided with the (intermittent) exclusive authorization to publish court reports and various official announcements, it is a unique source for both historians and linguists.

Based on facsimiles provided by the Austrian National Library’s digitization program ANNO, a workflow has been developed in order to edit the newspaper using the TEI Guidelines. Selected issues from different decades of the 18th century have been transcribed with Transkribus, a platform originally developed to support handwritten text recognition. Subsequently, the texts have been enriched with structural and semantic markup. Here, our focus lies in defining (preliminary) models, addressing questions like: How can the semantics of “news sections” be made accessible in a machine-readable manner? How are the various types of lists (births, deaths, arrivals of important persons) to be encoded, as they develop over the course of time?

The sample texts of this pilot study have been linguistically annotated with part-of-speech and lemma information using the TokenEditor, a web-based interface for collaborative annotation. As the tool was used before on texts from a similar historic period (namely for tagging the Austrian Baroque Corpus), it proved well trained for this task.

By describing the applied workflow, the poster will discuss period-specific challenges and limitations of these early modern textual resources and highlight potential benefits for various disciplines.

References

- Resch, Claudia; Czeitschner, Ulrike (eds.) (2015): ABaC:us – Austrian Baroque Corpus. <https://acdh.oeaw.ac.at/abacus/> [2016-07-31].
- ANNO – AustriaN Newspapers Online. <http://anno.onb.ac.at/> [2016-07-31].
- Burger, Harald; Luginbühl, Martin (2015): *Mediensprache*. Berlin / Boston. 39-45.
- Reisner, Andrea; Schiemer, Alfred (2016): Das Wien(n)erische Diarium und die Entstehung der periodischen Presse. In *Österreichische Mediengeschichte 1*. Wiesbaden. 87-112.
- Resch, Claudia; Czeitschner, Ulrike (2016): Morphosyntaktische Annotation historischer deutscher Texte: Das Austrian Baroque Corpus. In: *Digitale Methoden der Korpusforschung in Österreich*. Reihe Linguistik und Kommunikationsforschung. Vienna.
- Schröder, Thomas (1995): *Die ersten Zeitungen*. Tübingen.
- Seitschek, Stefan (2013): Religiöse Praxis am Wiener Hof: Das Beispiel der medialen Berichterstattung. In *Frühneuzeitforschung in der Habsburgermonarchie*. Vienna. 71-102.

François Vignale

EuRED: XML-TEI in database structure

The EuRED (European Reading Experience Database) project¹ consists of a RDBMS gathering reading experiences made by Europeans from the 15th century until today. The database has been inspired by the purely British UK-RED database² which was launched in 2006 and contains approximately 40,000 records. Our goal, by achieving a proof of concept, is to give UK-RED a chronological and geographical extension with a European dimension, to take care both of new sources (CMCs, etc.) and new needs expressed by researchers. This is why new fields necessary to describe reading experiences have been added to UK-RED's data model. It was very quickly decided to reuse and convert the original data into TEI P5 and, therefore, to build a new data model also based on TEI P5. The choice of TEI P5 is relevant for two main reasons: First, around 90% of the UK-RED fields can easily be replaced by existing TEI P5 elements; second, TEI P5 is customizable. Thus a new class (model.experience-DescPart) containing 23 new elements mostly linked with specific thesaurii has been designed to help describing reading experiences. In EuRED's data model, these new elements are placed in the `<teiHeader/>` and not in the `<body/>`. Besides the challenge of the conversion of UK-RED's data, this decision was made because EuRED is not a database gathering textual sources containing transcriptions of reading experiences, but, on the contrary, a database gathering transcriptions of reading experiences located in textual sources. The aim of this poster is to explain and demonstrate how we use TEI P5, not only for encoding textual sources but also for data model building and database management purposes.

¹ <http://eured.univ-lemans.fr> [2016-07-31].

² <http://www.open.ac.uk/Arts/reading/UK> [2016-07.31].

Joseph Wang

From DOCX via TEI to Literature Map

Data in literary studies often have references to location and time: Writers do travel and stage scenes in their novels often in places they have visited. By conveying these data together with geographical maps, we are able to create a so-called literature map that gives us an overview of the relationship between biography and writing of an author. But we can also compare depictions of a place across several authors, which is of interest for historians and geographers. Of course, these data is also useful for literary traveler's guides.

The project "Tyrol / South Tyrol – A literary topography" (funded by Austrian Science Fund FWF, P26039) aims at the gathering of such geo- and time-referenced data and the creation of literature maps for the Tyrolean region. We use a database consisting of two parts for our work: a document repository (for TEI data and images), and a RDBMS (managing metadata of authors, places, keywords, texts, dates, as well as relationships among them). Open government data and authority records are used for raw maps and metadata on places and people, TEI and DOCX are used for gathering literary data. By doing so, we can use a text-centric workflow by annotating texts, which is more natural for scholars in the humanities.

For each author and place, we create a DOCX document. A table at the beginning of the document captures metadata of the entities; biographical notes and primary texts follow the table. Using the stylesheet- and remark-function of MS Word, we are able to add annotations on places, dates and keywords to the texts. The data-ingest routine converts DOCX

to TEI, and then it extracts data from the TEI-document for the relational database. After the conversion has been reviewed by scholars, the data is then ingested to the document repository and the relational database.

Joseph C. Wicentowski, Magdalena Turska

Bringing TEI PM to Foggy Bottom

In May 2016, the U.S. State Department's Office of the Historian relaunched its public website, history.state.gov, and the new TEI Processing Model (TEI PM) played a key role in the site's plan for scalability and sustainability. By replacing custom-written code with TEI PM, the project shed years of legacy, custom-written code – laden with duplication and conditional branches for different publications and output formats – and replaced it with a light and lean ODD file containing a single set of TEI Processing Model instructions that form the basis of all transformations for the site's TEI-based publications: HTML, EPUB, and PDF.

This poster will illustrate key aspects of the architecture of the new site: It uses the TEI PM Toolbox, an implementation of TEI PM that generates XQuery 3.1-based stylesheets for multiple output formats from standard TEI PM ODD files. The poster will also illustrate how TEI PM helped the Office of the Historian achieve its goals for scalability and sustainability: improved performance under load, less custom code, and better separation of concerns among editors / designers / programmers by reducing reliance on programmers for stylesheet maintenance and empowering editors to define and improve how their TEI is processed. By documenting the concrete benefits of TEI PM (using measurable benchmarks where possible, such as lines of code), the poster's authors hope to inform other projects that might have similar goals about this significant new facility of the TEI.

References

history.state.gov: A case study of Digital Humanities in Government (2011). In Journal of the Chicago Colloquium on Digital Humanities and Computer Science 1/3. <https://letterpress.uchicago.edu/index.php/jdhcs/article/view/80> [2016-07-31].

Mary Erica Zimmer

**TEI in the Archives:
Unfolding the Bookshops in Paul's Cross Churchyard**

Writing in 2000 of “[d]igital archives combined with new technologies,” Seamus Ross lauded this union’s ability to support “simultaneous access to a range of sources” as well as development of “research methods not possible with [...] printed or hand written records” alone. Less frequently discussed, however, are ways digital connections may enrich understanding of print and manuscript records themselves, while encouraging scholarship that draws upon both. Recent projects such as Shakespeare Documented show digitized archives’ potential not only for reuniting fragmented artifacts, but also for evolving new, composite wholes to support research.

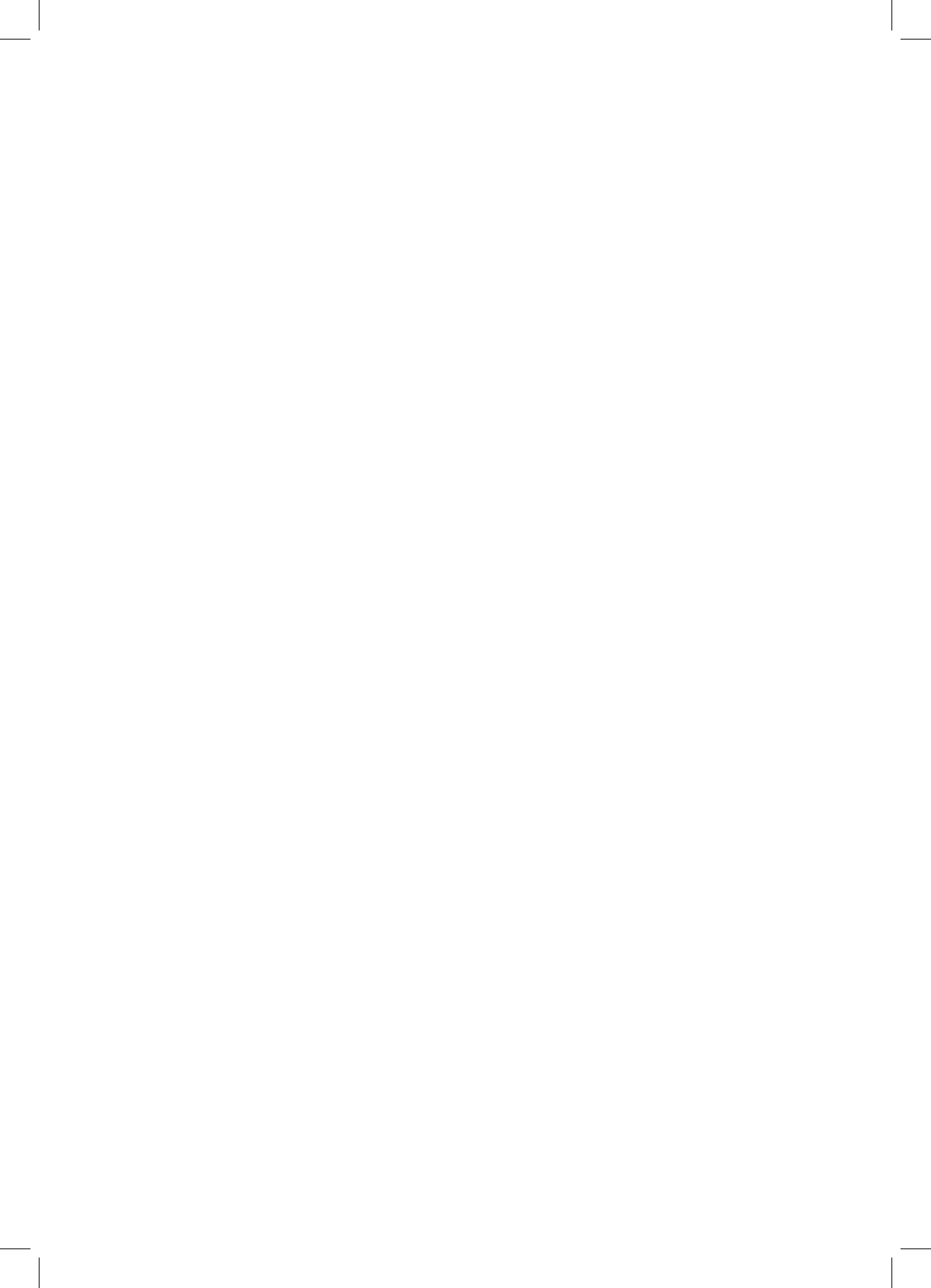
Such potential extends as well to encouraging engagement with existing scholarship. Among salient examples potentially complemented by an approach of this kind would be the landmark work of Peter W. M. Blayney, whose 1990 *The Bookshops in Paul's Cross Churchyard* uses meticulous archival research to map the evolving footprint of the shops and their inhabitants. Resulting has been a powerful print visualization: a composite, layered image of shops, their locations, and their inhabitants. Centrally, this project proposes to create a digitized model of the context Blayney has rendered, while connecting aspects of the reconstruction to the documents used to create it. Making the research’s archival foundations visible honors the initial, seminal achievement while encouraging exploration – and emulation – of the work.

The proposed project's TEI encoding of key documents will observe existing permissions while developing a composite resource. Connecting print and manuscript archives through a third medium will provide perspective on ways early modern records may present reciprocal and reinforcing information. Intersections with further early modern corpora also appear promising, particularly given the January 2015 release of over 25,000 TEI-compliant SGML-encoded EEBO-TCP texts. Ultimately, TEI's role as a common language remains vital to the project's success.



Demonstrations





Hugh Cayless, Raffaele Viglianti

CETELcean, a JavaScript library for isomorphic TEI to HTML transformation

This demonstration will introduce CETELcean, a JavaScript library for displaying TEI in a web browser. Instead of sticking with the semantically poor element set of HTML, CETELcean reframes an isomorphic transformation of TEI as HTML, by registering modified TEI elements with the browser using the new Custom Elements technology¹. This is a new feature that does not yet work in all browsers, but CETELcean supports older browsers too.

Typical TEI workflows first mark up text and then transform it into a variety of formats for digital publication, which is particularly useful where multiple outputs are desired. Converting TEI to HTML is the most common and most practical way of publishing TEI texts on the web, but HTML lacks what TEI has: a very well-considered and mature set of semantic tags for encoding texts. When converting TEI to HTML, the semantic distinctions in the markup are often lost in favor of typographic distinctions in the display. In other words, the data model represented in the TEI fails to carry over to the online version.

The solution is to preserve the semantics of the TEI source. Various ways of doing this have been proposed. These include the imposition of TEI semantics onto HTML elements using RDFa or microformat, and (more practically) the use of an in-browser XSLT transformation to wrap the TEI document in an HTML envelope and make it able to be styled with

¹ <https://w3c.github.io/webcomponents/spec/custom/> [2016-07-31].

CSS and JavaScript². CETELcean's approach is inspired by Boilerplate, but it is based on web standards and is more flexible. It does not require an in-browser XSLT step, nor any modification to the source XML; TEI content can be loaded in the browser via an AJAX call, or via server-side processing. A Boilerplate-style XSLT transform directive could be used too.

² See TEI Boilerplate. <http://dcl.ils.indiana.edu/teibp/> [2016-07-31].

André De Tienne

**STEP Transcriptor:
a comprehensive TEI-XML compliant transcription app**

STEP Transcriptor is the most advanced of several tools being developed to accompany the creation of the online Scholarly Text Editing Platform, a general solution that provides a complete transcription, editing, annotation, and layout workflow for scholarly editing projects. STEP Transcriptor has been designed and engineered to help transcribers encode complex texts easily while complying with TEI's manuscript transcription encoding rules. The app (for Mac and Windows) takes care of nearly all of the tedious aspects of TEI encoding. It provides five distinct encoding methods (via pull-down menus, pop-up menus, function keys, keyboard shortcuts, and specialized tagging cursors) that help insert tags, attributes, and values accurately while minimizing typing. Several specialized widgets automatically take care of especially complex encoding situations (including nested alterations and transpositions). STEP Transcriptor automatically creates lists of well-described alterations. It also provides a separate field to enter all sorts of descriptive tags, separately from transcriptive tags. Everything about the app is customizable: the content of menus, the selection of TEI tags, attributes, and values, even tag colors. The software can be used to start transcriptions from scratch or to improve existing transcriptions, all of which can be imported or exported in a variety of file formats. It can be used independently of the STEP platform. A key objective is to help editorial projects train and retain transcribers by providing them with a sophisticated tool that is reliable and easy to use. The tool helps flatten TEI's learning curve while producing correctly encoded transcriptions. It provides multiple

views of the texts, rendered or not rendered, encoded with or without anchors. The demonstration will focus on the core functionalities of the application.

Eugene W. Lyman II

The Elwood Markup & Text Analyzer (TEMTA)

I will demonstrate interactive software that combines a number of important functions to assist scholars who wish to make an in-depth assessment of unfamiliar TEI encoded texts – an increasingly frequent circumstance at a time when the comparability of encoding practices and the resulting interoperability of marked-up texts has become the object of special concern.

TEMTA is designed

1. to produce sophisticated views of markup density and hierarchical structure as well as targeted inventories of the elements, attributes, and values of any TEI encoded text, and
2. to facilitate sophisticated, near-instantaneous regular expression searching of text, markup, and varied combinations of the two.

It thus provides the basis for nuanced, iterative investigations in cases where the unknowns associated with a text's markup and lexical content are high with the consequence that a step-wise refinement of scholarly inquiry is very much in order. TEMTA includes a built-in capacity to graph the quantity and distribution of found features across a variety of appropriate textual or bibliographic units (chapters, folios, etc.) as well as the ability to export JSON and other encoded files that can be cast as visualizations by standard graphing software. Its handling of search results includes, among other options, the highlighting of found features in lists which in turn function as clickable indices to the display of each feature within its original context.

TEMTA's design and functionality have been based on my own wishes to investigate texts and their markup in ways that will both take the measure of the scholarship that has gone into their encoding as well as attain new insight into the nature of the stylistic and structural features that have thereby been made accessible to digital inquiry.

Magdalena Turska, Wolfgang Meier

Early Print: annotation module for TEI

EarlyPrint project aims to create an environment for collaborative curation of EEBO collection, spanning over 60 thousands of texts. Its core part and subject of this demonstration, currently under development, is an annotation module tightly integrated with the reading view. The process of annotation is thus highly intuitive, resembling the workflow with paper and pencil, which encourages work on the text in a natural way while reading it.

Annotations, which can be roughly classified into two categories:

- editorial changes or text critical annotations
- public or private notes which comment on the text and may eventually lead to discussion threads developing as other users refer to existing comments

will be all stored in a stand-off format aligned with W3C's Web Annotation Data Model. The first category of annotations is supposed to result, after annotation review process, in a new version of established TEI encoding which merges in accepted annotations. To achieve this goal, the project employs a technology called web components, which allows to define custom HTML elements and attach actions, as well as user interface components to them. Combined with use of TEI Processing Model for transformations into HTML enriched with custom elements, this approach leads to a modular design, allowing to add more components for different types of editing tasks over time. The project's timeline assumes starting with a small number of components to handle the

most common defects in EEBO, and later adding more without having to change the core application.

Exposing annotations using the Web Annotation Data Model, we aim to foster interoperability with other, similar software, and add another layer of abstraction to make it easier to build upon and extend for purposes which are beyond the immediate scope of the EEBO project.



Workshops





Piotr Bański, Karlheinz Mörth, Laurent Romary, Andreas Witt

Lexical information in the TEI: modeling and markup

Quite a while has passed since the rewrite of the chapter on Dictionaries in the TEI Guidelines, and many projects have since then adopted the new proposals. They have also generated some new discussion (e.g. in the case of the generalized usage of `<cit>` and `<quote>`) and are still in the process of getting refined (with the grammatical information section recently having become more compact). Requests for new features and improvement are continually processed by the TEI Council and discussed by the community in the GitHub tickets.

In the meantime, Subcommittee 4 of the ISO Technical Committee 37 has published its recommendations for the encoding of lexical databases (ISO LMF), accompanied by a serialization that hasn't become widespread in the field. The Subcommittee is in the process of revising the LMF, with an eye to both improvements in the model and to its TEI serialization, which, subject to the TEI Council's approval, might potentially yield a new version of the "Dictionaries" chapter.

The workshop, organized as an extension of the Linguistics SIG meeting, will report on the history of lexical description in the TEI, the current state of the art, and the envisioned developments. An essential projected part of the workshop will be a discussion of the currently hottest topics (among others, the emerging serialization of ISO LMF in the TEI, its envisioned coverage, feature set, and the timeline / milestones for the expert / developer group).

Stefan E. Funk, Thomas Kollatz, Oliver Schmid, Sibylle Söring, Ubbo Veentjer

Around the world in a TextGrid day. Edit, manage, publish, and explore your TEI data with TextGrid and DARIAH-DE

This workshop can be seen as a follow-up to the TEI 2015 Lyon panel on “Shared Platforms for TEI Community: TextGrid, TAPAS, DARIAH” which discussed various TEI based repositories, common user requirements and research scenarios, and their technological frameworks.

The workshop will explore two infrastructures, TextGrid and DARIAH-DE, at work, focusing on modeling a “prototype” digital edition starting with a plain text, finishing with a TEI-based online publication.

Outline

Digital editions are one of the core applications of TEI. However, when it comes to not only generating, but publishing enriched data in a standardized, interoperable way, researchers find themselves confronted with a variety of both scholarly and technological decisions and requirements. TextGrid and DARIAH-DE support researchers during the entire process of generating, archiving, and presenting their TEI data in an interoperable environment.

TextGrid is a virtual research environment for the Humanities funded by the German Ministry for Education and Research (BMBF). It enables the collaboration of researchers at different places as well as archiving and publication of research data in an XML-based repository, following international standards. Core components are the TextGrid Repository, an XML-based long-term research data archive, and the TextGrid

Laboratory, an adaptable software bundle for collaborative digital scholarly editing based on XML/TEI. TextGrid is part of the digital research infrastructure DARIAH-DE, equally funded by the BMBF, to provide a sustainable digital infrastructure for humanities research. Users benefit from both portfolios with one single login.

On the basis of text extracts from Jules Verne's "Around the World in 80 Days", participants will transcribe, markup, annotate and geo-refer TEI data with the help of TextGridLab and the DARIAH-DE Geo-Browser. The encoded data will then be published via the TextGrid and DARIAH-DE repository and into a web platform (online edition) through SADE (Scalable Architecture for Digital Editions).

Workshop participants will explore the diversity of the digital research environment TextGrid as well as tools and services of the research infrastructure DARIAH-DE and its possibilities for digital analysis. An introduction to both infrastructures, illustrated by examples from various user scenarios, will be followed by a hands-on session focusing on editing and modeling TEI data, particularly of digital editions, to their publication and exploration. Following these steps, participants will be able to explore the different tools and services to generate, model and publish XML/TEI data guided by TextGrid and DARIAH staff members in small groups according to their own interests and needs. The aim is to create a small digital edition starting with a facsimile and a transcription, enriching data with extended markup such as geospatial encoding, and publishing into both the TextGrid Repository and a SADE instance (web portal).

step 1 – Users and roles

- The Project and User Management to administrate folders, files and access rights.

step 2 – Editing

- An XML editor allows working with XML files by offering the user a tree-, source-code, WYSIWYM and web preview. A Unicode Character Table allows easy searching, copying and inserting of any symbol in the Unicode character set. Alternatively, oXygen XML Editor can be embedded.
- A customizable Metadata Editor for creating and administrating metadata of TextGrid objects that can be used for project-overlapping queries.
- The Text Image Link Editor supports the XML Editor in linking text and image segments as well as serving as an image annotator. An output file links text elements with the topographical position of a rectangular or polygonal image segment as SVG, as needed for linking facsimiles and transcriptions in digital editions.
- The image viewer and referencing tool DigiLib covers a gallery view of several images, zoom, scaling, marking, and referencing functions.

step 3 – Publishing

- Publishing into the TextGrid Repository: A digital long-term archive for humanities research data to guarantee long-range availability and accessibility together with extensive search facilities, different download formats, and tools for further exploration and visualization.
- The software SADE – Scalable Architecture for Digital Editions supports the publication of TEI data in the TextGridLab into web portals.

step 4 – Exploring published editions in the repository

- Browsing and exploring published editions within TextGridRep and SADE, including persistent identifiers for referencing each object.
- Visualization of spatio-temporal information via the DARIAH-DE Geo-Browser.
- Reference tools for both text and image annotation, creating interoperable and referenceable data.
- Visualization of quantitative text analyses with DigiVoy.

References

- DARIAH-DE. <https://de.dariah.eu/dariah-de-english> [2016-07-31].
- TextGrid. <https://textgrid.de/en/> [2016-07-31].
- TextGrid Registration. <https://textgrid.de/registrierung/download> [2016-07-31].
- TextGrid Lab (Download). <https://textgrid.de/download> [2016-07-31].
- TextGrid Rep. <https://textgridrep.org> [2016-07-31].
- TextGrid Online Documentation and Wiki. <https://wiki.de.dariah.eu/display/TextGrid/Main+Page> [2016-07-31].
- DARIAH-DE Geo-Browser and Datasheet Editor. <http://geobrowser.de.dariah.eu> [2016-07-31].
- DigiVoy. <https://de.dariah.eu/digivoy> [2016-07-31].
- SADE. <http://sade.textgrid.de/exist/apps/SADE/textgrid/index.html> [2016-07-31].
- Blümm, Mirjam; Funk, Stefan E.; Söring, Sibylle (2015): Die Infrastruktur-Angebote von DARIAH-DE und TextGrid. In *Information. Wissenschaft & Praxis* 66/5-6. 304-312. <http://dx.doi.org/10.1515/iwp-2015-0055> [2016-07-31].
- Hedges, Mark; Neuroth, Heike; Smith, Kathleen M.; Blanke, Tobias; Romary, Laurent; Küster, Marc; Illingworth, Malcolm (2013): TextGrid,

TEXTvire, and DARIAH: Sustainability of Infrastructures for Textual Scholarship. In Journal of the Text Encoding Initiative 5. [https://textgrid.de/fileadmin/publikationen/TextGrid_TEXTvire_and_DARIAH - Sustainability of Infrastructures for Textual Scholarship_Hedges_Neuroth_Smith_Blanke_Romary_K%C3%BCster_Illingworth_2013.pdf](https://textgrid.de/fileadmin/publikationen/TextGrid_TEXTvire_and_DARIAH_-_Sustainability_of_Infrastructures_for_Textual_Scholarship_Hedges_Neuroth_Smith_Blanke_Romary_K%C3%BCster_Illingworth_2013.pdf) [2016-07-31].

Kollatz, Thomas (2016): Raum-Zeit-Analysen mit Geo-Browser und Data-sheet-Editor. In Bibliothek – Forschung und Praxis 40/2/1-6. Neuroth, Heike et al. (eds.): Sonderheft DARIAH-DE. 229-233.

Neuroth, Heike; Rapp, Andrea; Söring, Sibylle (eds.) (2015): TextGrid: Von der Community für die Community – Eine Virtuelle Forschungsumgebung für die Geisteswissenschaften. http://www.univerlag.uni-goettingen.de/handle/3/Neuroth_TextGrid [2016-07-31].

Söring, Sibylle (2016): Technische und infrastrukturelle Lösungen für digitale Editionen: DARIAH-DE und TextGrid. In Bibliothek – Forschung und Praxis 40/2/1-6. Neuroth, Heike et al. (eds.): Sonderheft DARIAH-DE. 207-212.

Martina Scholger, Peter Stadler

TEI2German Translatathon

The workshop is an open call for the German-speaking community to improve the TEI by providing concise and fitting translations of the descriptive texts of the TEI P5 guidelines' specs. The goal of the workshop is to translate `<desc>` and `<gloss>` elements and feed these back into the TEI specifications. The setting will be hands-on and interactive. This event will be specifically targeted at the German-speaking community, which will most likely be a substantial part of the audience at Vienna, with iterations for other languages envisioned at subsequent TEI events. We aim to add and improve translations in areas that are identified as the first session of the translatathon, so that it might become possible to have coherent translations for an identified set of tags. This will be supported by supplemental material that groups the existing tags by module and / or area of interest, to make the progress visible.

Magdalena Turska, Lars Windauer

eXistdb: more than a database

Day 1: Crash course in eXistdb basics for TEI data

Day 2: X-wiZZardry: application building with eXistdb

A significant portion of humanities research data is marked up and stored as XML documents. In order to provide standard interactive features for digital research resources such as full text search possibilities, it is obvious to use a database which is optimized for the handling of XML documents. eXistdb is one of the leading native XML databases currently available. This open source solution is becoming a technology of choice in Digital Humanities, especially when it comes to storing, querying and publishing large corpora of XML marked-up texts. eXistdb though is much more than just a database: It is also an application platform, web server and even a document creation platform. eXistdb is also currently the only out-of-the-box implementation of the TEI Processing Model. Thus, it is extremely well suited to be a database, rapid application development and publishing solution of choice for large, established collaborative projects and individual researchers alike.

This tutorial aims for a mix of talks and hands-on practical sessions. It will be divided in two parts, lasting a day each, which can be taken independently. Part one will provide a basic introduction to eXistdb and follow on to querying data collections with XPath, to conclude with publishing TEI documents with the TEI Processing Model.

Part two will move beyond out-of-the-box solutions and will discuss XQuery, a language which is used to query, extract, and manipulate

XML documents as well as build applications for eXistdb. We will also introduce the principles of eXistdb indexing and profiling features. As a source of inspiration and a panorama of applications that could benefit from relying on eXistdb as its core technology, we will also showcase a few research projects using eXistdb.

As eXistdb is fully based upon Open Standards and Open Source, it is a future-proof and sustainable choice for research projects. Since its inception in 2001, eXistdb development has always been driven by the needs of a large user community. We believe that the TEI conference is a good opportunity for scholars to meet eXistdb contributors and practitioners, and to exchange ideas how the Digital Humanities community could both engage and benefit further from eXistdb development.

The introduction to the features of eXistdb during the workshop will give participants a starting point for their own eXistdb-based project as well as help them narrow their choices when it comes to selecting database technology suitable for their research.

Tanja Wissik

ACDH Tool Gallery 2.2: TEI Basics - A gentle introduction to Text Encoding

The ACDH Tool Gallery 2.2 “TEI Basics – A gentle introduction to Text Encoding” will take place in the context of the TEI 2016 conference. The event will introduce researchers from different disciplines who do text-based work in their research to the TEI markup language. The ACDH Tool Gallery will give an overview of the “tag set” of XML “elements” that are used to encode texts, along with “attributes” used to modify these elements and will show how the TEI standard is structured. After an introduction in the morning, the participants will get an insight into two TEI-based projects before trying to encode their own text material according to the TEI standards with the guidance of experienced scholars and practitioners in the field.

The detailed program can be found here
<http://acdh.oeaw.ac.at/acdh/en/acdh-tool-gallery-2.2>

Special Interest Group meetings

Computer-Mediated Communication SIG

Correspondence SIG

TEI for Linguists SIG

Text & Graphics SIG

Preparatory meeting: Towards a SIG Periodica



Index





Amminger, Agnes | Salzburg Mozarteum Foundation, Austria **89**

Andorfer, Peter | Austrian Centre for Digital Humanities, Austrian Academy of Sciences, Austria **121, 127, 133**

Andrews, Tara Lee | Department of History, University of Vienna, Austria **23**

Armaselu, Florentina | University of Luxembourg, Luxembourg **31**

Bański, Piotr | Institute for the German Language, Mannheim, Germany **35, 175**

Beck Mink, Meridith | Council of Library and Information Resources, Knox College, United States of America **67**

Beißwenger, Michael | Department of German Studies, University of Duisburg-Essen, Germany **39**

Bermudez Sabel, Helena | Departamento de Filoloxía Galega, Universidade de Santiago de Compostela, Spain **43**

Bia, Alejandro | Departamento de Estadística, Matemáticas e Informática, Universidad Miguel Hernández, Spain **45**

Bleier, Roman | Center for Information Modelling – Austrian Centre for Digital Humanities, University of Graz, Austria **49, 51**

Boenig, Matthias | Deutsches Textarchiv, Berlin-Brandenburg Academy of Sciences, Germany **69**

Börner, Ingo | Department of German Studies, University of Vienna, Austria **139**

Bowers, Jack Thompson | Austrian Centre for Digital Humanities, Austrian Academy of Sciences, Austria **53, 55**

Breen, Richard | Maynooth University, Ireland **49**

Breuil, Sandrine | Centre d'Études Supérieures de la Renaissance, Université François-Rabelais, Tours, France **111**

Broughton, William Michael | Cologne Center for eHumanities, University of Cologne, Germany **57**

Cano, Jesús | Universidad Nacional de Educación a Distancia, Spain **73**

Caminero, Agustín C. | Universidad Nacional de Educación a Distancia, Spain **73**

Cayless, Hugh | Duke University Libraries, Duke University, United States of America **163**

Chartrand, James | Open Sky Solutions, Canada **65**

Ciotti, Fabio | Dipartimento di Studi Filologici, Linguistici e Letterari, Università Roma Tor Vergata, Italy **59**

Cummings, James | IT Services, University of Oxford, United Kingdom **63, 65, 79**

Czmiel, Alexander | TELOTA, Berlin-Brandenburg Academy of Sciences, Germany **69**

Dalmau, Michelle | Indiana University Libraries, Indiana University, United States of America **67**

De Tienne, André | School of Liberal Arts, Indiana University at Indianapolis, United States of America **123, 165**

Declerck, Thierry | German Research Center for Artificial Intelligence, Germany **53, 99, 125**

Del Rio Riande, Gimena | Consejo Nacional de Investigaciones Científicas y Técnicas, Universidad de Buenos Aires, Argentina **73**

Duboc, Mathieu | Centre d'Etudes Supérieures de la Renaissance, Université François-Rabelais, Tours, France **111**

Dumont, Stefan | TELOTA, Berlin-Brandenburg Academy of Sciences and Humanities, Germany **69**

Ehrhardt, Eric | Department of German Studies, University of Mannheim, Germany **39**

Erjavec, Tomaž | Department of Knowledge Technologies, Jožef Stefan Institute, Slovenia **91**

Faghihi, Yasmin | Near and Middle Eastern Department, Cambridge University Library, United Kingdom **71**

Fields III, Archie | Department of History and Philosophy of Science and Medicine, Indiana University, United States of America **67**

Fischer, Nora | Institute for History of Art and Musicology, Austrian Academy of Sciences, Austria **149**

Franzini, Greta Hayley | Göttingen Centre for Digital Humanities, Georg-August-Universität Göttingen, Germany **127**

Funk, Stefan E. | Göttingen State and University Library, Germany **177**

Gaiffe, Bertrand | Analyse et Traitement Informatique de la Langue Française, Université de Lorraine, France **35**

Geißler, Nils | Cologne Center for eHumanities, Cologne University, Germany **101**

Gibson, Nathan P. | School of Divinity, Vanderbilt University, United States of America **129**

González-Blanco, Elena | Departamento de Literatura Española y Teoría de la Literatura / Laboratorio de Innovación en Humanidades Digitales, Universidad Nacional de Educación a Distancia, Spain **73**

Haaf, Susanne | Deutsches Textarchiv, Berlin-Brandenburg Academy of Sciences and Humanities, Germany **69**

Hadden, Richard | Maynooth University, Ireland **51**

Hannessschläger, Vanessa | Austrian Centre for Digital Humanities, Austrian Academy of Sciences, Austria **77, 149**

Hebenstreit, Desiree | Department of German Studies, University of Vienna, Austria **131**

Herberth, Arno | Department of German Studies, University of Vienna, Austria **131**

Hernández, Roberto | Escuela Técnica Superior de Ingeniería Informática, Universidad Nacional de Educación a Distancia, Spain **73**

Herold, Axel | CLARIN-D / Kyoto, Berlin-Brandenburg Academy of Sciences, Germany **39**

Holford, Matthew Lawrence | Bodleian Library, University of Oxford, United Kingdom **79**

Hooper, Wallace | Department of the History and Philosophy of Science, Indiana University, United States of America **67**

Jocham, Tobias J. | Histoire et Littérature, Collège de France / Berlin-Brandenburg Academy of Sciences, France **71**

Jones, Huw | Cambridge University Library, United Kingdom **79**

Kampkaspar, Dario | Herzog August Bibliothek Wolfenbüttel, Germany **121**

King, Margaret | Austrian Centre for Digital Humanities, Austrian Academy of Sciences, Austria **133**

Klugseder, Robert | Institute for History of Art and Musicology, Austrian Academy of Sciences, Austria **81**

Kodolányi, Judit | Petőfi Literary Museum, Hungary **83**

Kollatz, Thomas | Salomon Ludwig Steinheim-Institut für deutsch-jüdische Geschichte, Germany **177**

Kovalenko, Kira | Institute for Linguistic Studies, Russian Academy of Sciences, Russia **99**

Kraft, Tobias | Berlin-Brandenburg Academy of Sciences and Humanities, Germany **69**

Lembi, Gaia | Department of Judaic Studies, Brown University,
United States of America **135**

Looschen, Sünna | Institute of Slavic Studies, Carl von Ossietzky
University, Germany **139**

Lopez, Patrice | Science Miner, France **35**

Lüngen, Harald | Institute for the German Language, Mannheim,
Germany **39**

Lyman II, Eugene W. | n / a, United States of America **167**

Mader, Anna | Institute for History of Art and Musicology, Austrian
Academy of Sciences, Austria **149**

Martínez, Clara I. | Departamento de Literatura Española y Teoría
de la Literatura, Universidad Nacional de Educación a Distancia,
Spain **73**

McAllister, Patrick | Institute for the Cultural and Intellectual
History of Asia, Austrian Academy of Sciences, Austria **85**

Meier, Wolfgang | eXist solutions, Germany **109, 169**

Meoni, Simon | Analyse et Traitement Informatique de la Langue
Française, France **35**

Michelson, David A. | Department of Classical Studies, Vanderbilt University, United States of America **129**

Mörth, Karlheinz | Austrian Centre for Digital Humanities, Austrian Academy of Sciences, Austria **87, 125, 175**

Morgenstern, Anja | Salzburg Mozarteum Foundation, Austria **89**

Mozhaev, Evgeny | Higher School of Economics Moscow, Russia **107**

Mueller, Martin | Department of English, Northwestern University, United States of America **19**

Muller, A. Charles | Center for Evolving Humanities, University of Tokyo, Japan **141**

Mylonas, Elli | University Library Center for Digital Scholarship, Brown University, United States of America **135**

Nagasaki, Kiyonori | International Institute for Digital Humanities, University of Tokyo, Japan **141**

Newman, William R. | Department of History and Philosophy of Science and Medicine, Indiana University, United States of America **67**

Ogrin, Matija | Institute of Slovenian Literature and Literary Studies,
Research Centre of Slovenian Academy of Sciences and Arts,
Slovenia **91**

Palmer, Sara | Emory Center for Digital Scholarship, Emory University,
United States of America **145**

Pastor, Rafael | Departamento de Sistemas de Comunicación y
Control, Universidad Nacional de Educación a Distancia, Spain **73**

Perathoner, Marcello | Cologne Center for eHumanities, Cologne
University, Germany **101**

Petrus, Alexander | Institut für Bibliotheks- und Informationswissenschaft,
Humboldt-Universität zu Berlin, Germany **117**

Pinche, Ariane | Centre Interuniversitaire d'Histoire et d'Archéologie
Médiévales, France **145**

Porte, Guillaume | Centre d'Etudes Supérieures de la Renaissance,
Université François-Rabelais, Tours, France **111**

Prag, Jonathan | Classics Faculty, University of Oxford, United Kingdom **65**

Pytlik Zillig, Brian L. | Centre for Digital Research in the Humanities,
University of Nebraska-Lincoln, United States of America **93**

Resch, Claudia | Austrian Centre for Digital Humanities, Austrian Academy of Sciences, Austria **19, 95, 149**

Robles-Gómez, Antonio | Departamento de Sistemas de Comunicación y Control, Universidad Nacional de Educación a Distancia, Spain **73**

Rogers, Chris | Bodleian Library, University of Oxford, United Kingdom **97**

Romary, Laurent | Institut national de recherche dédié au numérique, France **35, 175**

Ros, Salvador | Escuela Técnica Superior de Ingeniería Informática, Universidad Nacional de Educación a Distancia, Spain **73**

Rosenblum, Brian | University of Kansas Libraries, University of Kansas, United States of America **133**

Satlow, Michael | Department of Judaic Studies, Brown University, United States of America **135**

Schmid, Oliver | Fachbereich Chemie, Technische Universität Darmstadt, Germany **177**

Schmidt, Thomas | Institute for the German Language, Mannheim, Germany **35**

Scholger, Martina | Center for Information Modelling – Austrian Centre for Digital Humanities, University of Graz, Austria **183**

Schopper, Daniel | Austrian Centre for Digital Humanities, Austrian Academy of Sciences, Austria **87, 95, 99, 149**

Schulz, Daniela Monika | Historisches Institut, University of Cologne, Germany **101**

Seltmann, Melanie | Austrian Centre for Digital Humanities, Austrian Academy of Sciences, Austria **55**

Shimoda, Masahiro | Center for Evolving Humanities, University of Tokyo, Japan **141**

Sievers, Martin | Trier Center for Digital Humanities, Trier University, Germany **103**

Skorinkin, Daniil | Higher School of Economics Moscow, Russia **107**

Söring, Sibylle | Göttingen State and University Library, Germany **177**

Stadler, Peter | Musikwissenschaftliches Seminar Detmold / Paderborn, University of Paderborn, Germany **35, 183**

Steiner, Christian | Institute of Economics, University of Graz, Austria **81**

Storrer, Angelika | Department of German Studies, University of Mannheim, Germany **39**

Tezarek, Laura | Literary Archives, Austrian National Library, Austria **131**

Thomas, Christian | Deutsches Textarchiv, Berlin-Brandenburg Academy of Sciences, Germany **69**

Tsuda, Tetsuei | National Research Institute for Cultural Properties, Japan **141**

Turska, Magdalena | IT-Services, University of Oxford, United Kingdom **109, 155, 169, 185**

Uetani, Toshinori | Centre d'Études Supérieures de la Renaissance, Centre National de la Recherche Scientifique, France **111**

Urizar, Miguel | Universidad Nacional de Educación a Distancia, Spain **73**

Veentjer, Ubbo | Göttingen State and University Library, Germany **177**

Viglianti, Raffaele | Maryland Institute for Technology in the Humanities, University of Maryland, United States of America **163**

Vignale, François | Centre de Recherches Historiques de l'Ouest – Le Mans, Université du Maine, France **151**

Voelkel, James | Beckman Center for the History of Chemistry, Chemical Heritage Foundation, United States of America **67**

Vogeler, Georg | Center for Information Modelling – Austrian Centre for Digital Humanities, University of Graz, Austria **113**

Walsh, John A. | School of Informatics and Computing, Indiana University, United States of America **67**

Wandl-Vogt, Eveline | Austrian Centre for Digital Humanities, Austrian Academy of Sciences, Austria **99**

Wang, Joseph | Institute of History and European Ethnology, University of Innsbruck, Austria **153**

Wicentowski, Joseph C. | U.S. Department of State, United States of America **155**

Windauer, Lars | eXist Solutions, Germany **185**

Wissik, Tanja | Austrian Centre for Digital Humanities, Austrian Academy of Sciences, Austria **187**

Witt, Andreas | Institute for the German Language, Mannheim, Germany **35, 175**

Wohlfarter, Eva | Austrian Centre for Digital Humanities, Austrian Academy of Sciences, Austria **149**

Zielke, Dennis | Computer and Media Service, Humboldt-Universität zu Berlin, Germany **117**

Zimmer, Mary Erica | Arts & Sciences Writing Program, Boston University, United States of America **157**

Zolles, Christian | Department of German Studies, University of Vienna, Austria **131**



Printed in Austria | September 2016

Austrian Centre for Digital Humanities, Sonnenfelsgasse 19, 1010 Vienna
www.oeaw.ac.at/acdh

