An introduction to encoding and processing text with TEI

WORKSHOP DIRECTORY: http://bit.ly/TEIWorkshop_Matls

SCHEDULE: http://bit.ly/ADH02018_TEIWorkshop

In her keynote speech at the Text Encoding Initiative (TEI) Conference in November 2017, Gimena del Rio Riande addressed a need for intensive training in TEI accessible to new scholars in the Latin American world. The TEI Technical Council wishes to respond to her call at ADHO 2018 in Mexico City with an intensive one-day orientation workshop to the TEI that offers hands-on training to participants in how to work with TEI for anyone new to XML and angle-brackets.

This workshop offers an exploration of scholarly text encoding, aimed at an audience of humanities students and scholars and new adopters of digital humanities. The TEI provides a widely used standard for the representation of texts in digital form and has become a fundamental technology for digital scholarship. In this workshop, we will devote special attention to the contexts and varieties of TEI encoding in digital projects, and we will explore how to write as well as how to read the TEI with a short, complementary introduction to XPath navigation. By the end of the day, our participants will have learned how to begin a TEI project and have gained skills in writing code, publishing it for presentation, and navigating it to read and comprehend TEI projects and gain ideas for their own work. The additional look at XPath navigation is designed as a gateway to learning how to process more than just editions with TEI. From this, participants will learn how to isolate patterns and extract data for analysis using simple XPath patterns, and they will gain a sense of *what* they can learn and *where* and *how* they can best learn it for further processing work.

Members of the TEI Council will conduct the training and also respond to the needs and questions of participants on site. The goal is to help new users learn basic skills that they can use to launch their own TEI projects and continue learning by themselves afterwards. Registrants will gain access to the oXygen XML Editor (with a complementary license key for 120 days), but will also be shown free editors and XML ecosystems (including TAPAS and eXist-db) that do not require purchasing a license. Participants will be introduced to a variety of tools and environments in which to work with TEI, and will also gain a free membership of the TEI community for one year following the workshop.

The morning session of this one-day workshop provides an introduction to text encoding with TEI as a valuable instrument in humanities research. The first part introduces the concept of markup as a way of making knowledge explicit and machine-readable, demonstrates the syntax and rules of XML (Extensible Markup Language), and provides a basic overview of the TEI Guidelines and some background information on the activities of this community-driven initiative. In preparing our first XML document, basic functionalities of the

oXygen XML Editor as an environment for encoding and querying XML-based documents will be introduced.
The second part of the morning session introduces the default TEI structure and some of the core components for getting started with the Text Encoding Initiative. Sample documents (e.g. a poem) for the hands-on-session will be provided in multiple languages.

The afternoon session will begin with a hands-on orientation to XPath with two purposes:

1) to learn how to navigate XML documents and locate information precisely, working with XML document structures. Participants will gain familiarity with XPath axes and predicates, and if time, basic XPath functions like distinct-values().
2) to use XPath to illuminate different structures of TEI, with examples of standard customizations such as EpiDoc, TEI for manuscripts, and others. XPath is taught as a way to learn to read the TEI.

The afternoon will continue with a preview of tools developed for the presentation and publication of TEI documents. This will introduce the TEI Stylesheets, CETEIcean, TEI Publisher, TAPAS, etc.
If there is sufficient time, we will demonstrate how to write simple XQuery scripts to extract lists of data in HTML and CSV formats.

At the conclusion of the workshop, participants will be informed about further possibilities and resources for continuing their TEI training, e.g. TEI ODD (make a schema to customize the TEI), practice and resources for further learning on your own, and locations for training. Furthermore, we want to discuss their user needs and solicit feedback, for example on employing the TEI specifications in other languages than English. Through a combination of hands-on practice, presentation and discussion, participants will learn the essentials of TEI markup and querying documents, leading them to realize how markup languages can support and enhance scholarship in the digital age.

Schedule

Part 1 (morning session): Introduction to encoding with TEI

- Introduction to Text Encoding and the TEI community
    - Markup, XML, TEI
    - Community resources
    - TEI Guidelines, Listserv, Wiki
- Introduction to XML
    - Elements and attributes; Tree hierarchy and its difficulties
    - Making your own XML vs. TEI standard
    - Hands-on session
- Environments for coding TEI
    - oXygen
    - atom.io
    - eXist-db
- TEI Basics
    - Basic Structure

    - Core Components
    - Hands-on session with sample documents

Part 2 (afternoon session): Introductory Hands-on Orientation to Reading and Navigating TEI with XPath and publishing your TEI files

- Work with either the free eXist-db installation or the oXygen XPath window: basic orientation skills with XPath axes
- XPath predicate "filters": How to isolate specific portions of documents
- Use XPath to survey and discuss project methods and structures for various kinds of TEI documents:
    - Poetry / Drama example(s)
    - Manuscripts example
    - EpiDoc example
    - XPath over a collection of files
- Basic processing with XPath
    - XSLT Identity transformation (a simple recipe to add line numbers)
    - XSLT or XQuery (in eXist-db): Extract a list of values from a regularly encoded collection of documents.
- Processing and publishing TEI documents: a preview
    - Using TEI Stylesheets
    - CETEIcean
    - TEI Publisher
    - TAPAS
- Conclusion and Feedback
    - Community issues: multilingualism in TEI
    - Further training opportunities
    - Participants' needs and discussion

Participation and requirements

The workshop is limited to a number of 25 participants. Past experience has shown that constant supervision by the instructors is needed to guarantee a high-quality workshop. No previous experience with XML, TEI and query languages is assumed. Participants should bring their own notebook and will be informed about required software prior to the workshop. There is no additional charge for workshop participants.

Bibliography

Beshero-Bondar, Elisa, *Newtfire training materials*, 2018. http://dh.newtfire.org.

Birnbaum, David, *<oo>→<dh> Digital humanities*. http://www.dh.obdurodon.org/.

Burnard, Lou, *What is the Text Encoding Initiative? How to add intelligent markup to digital resources*. OpenEditionPress, 2014. http://books.openedition.org/oep/426

CETEIcean. TEI in HTML5 custom elements. https://github.com/TEIC/CETEIcean.

Flanders, Julia, Syd Bauman, and Sarah Connell. “XSLT.” and “Text Encoding with TEI.” In *Doing Digital Humanities: Practice, Training, and Research*, ed. Constance Crompton, Ray Siemens, and Richard Lane. Routledge Press, 2016.

Flanders, Julia and Fotis Jannidis. “Data Modeling.” In *A Companion to Digital Humanities*, ed. Susan Schreibman, Ray Siemens, and John Unsworth. Wiley-Blackwell, 2016.

Hawkins, Kevin S. *Introduction to XML for Text*, 2015. http://www.ultraslavonic.info/intro-to-xml/.

Kay, Michael. *XSLT 2.0 and XPath 2.0 Programmer's Reference*. Wrox, 2008.

Tapas Project. Visualize, Store, and Share Your TEI. http://tapasproject.org/

TEI Consortium, eds. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. 3.3.0. 2018-01-31. TEI Consortium. http://www.tei-c.org/Guidelines/P5/.

TEI Publisher. The Instant Publishing Toolbox. https://teipublisher.com.

Workshop Instructors

Elisa Beshero-Bondar, PhD
Director, Center for the Digital Text and Assoc. Prof. of English, University of Pittsburgh at Greensburg

An active member of the Text Encoding Initiative (TEI), Dr. Beshero-Bondar was elected to serve from 2016 - 2019 on the TEI Technical Council, an eleven-member international committee that supervises amendments to the TEI Guidelines. A faculty member at Pitt-Greensburg since 2004, she serves as the Director of Pitt-Greensburg's Center for the Digital Text. Dr. Beshero-Bondar researches British Romanticism in poetry and drama from the 1790s - 1830s. Her book about women Romantic poets, titled *Women, Epic, and Transition in British Romanticism*, was published by the University of Delaware Press in 2011. Her published articles in ELH (English Literary History), Genre, Philological Quarterly, and The Wordsworth Circle investigate the poetry of Robert Southey, Mary Russell Mitford, and Lord Byron in context with 18th- and 19th-century views of revolution, world empires, natural sciences, and theater productions. Her research often involves working with archival letters and manuscripts, and has led her to study what early 19th-century poets and dramatists understood about human physiology and electricity, cultural first contact on Pacific islands, and the mutiny on the HMS Bounty. She is the architect of the Digital Mitford Project and other digital research projects involving TEI XML to build editions and prepare structured analyses of variants and collocations in texts. Find her and her projects on GitHub at https://github.com/ebeshero.

Elli Mylonas
Director, Center for Digital Scholarship, Brown University Library
Elli Mylonas has been working on the Text Encoding Initiative since its early days, when she was a member of several working groups and she is currently a member of the TEI Technical Council. She provides technical consultation and support for two epigraphic corpora that use the Epidoc customization of the TEI, and has taught both TEI and Epidoc workshops. She has been supporting and developing different sorts of digital humanities projects at Brown University since 1994, first at the Scholarly Technology Group, and now in the University Library's Center for Digital Scholarship. She was managing editor of the Perseus Project in its infancy. Her background is in Classics.

Martina Scholger
Centre for Information Modeling - Austrian Centre for Digital Humanities
University of Graz
Martina Scholger is a research associate at the Centre for Information Modelling – Austrian Centre for Digital Humanities at the University of Graz. She studied Art History and is currently preparing her dissertation on the potential of digital editions in the analysis and reconstruction of artistic association processes. In addition to teaching data modelling, text encoding and X-technologies, her work at the centre involves the conceptual design, development and implementation of numerous cooperation projects in the field of digital humanities (see http://gams.uni-graz.at). She has been a member of the Institute for Documentology and Scholarly Editing since 2014 and an elected member of the TEI Technical Council since 2016, where she is currently its Chair. She has been teaching at a number of Summer Schools in the context of digital scholarly editing, e.g. at the Digital Humanities at Oxford Summer School and Schools organised by the Institute for Documentology and Scholarly Editing (IDE).

NOTES:

Elisa: Arrangements with oXygen ASAP [done]

Prepping files we know we want to work with
--available online [done]
-- AND thumbdrive(s) [NOT done]

PUBLIC SCHEDULE: http://bit.ly/ADH02018_TEIWorkshop

9-10:30

Introductions (15 minutes?) (all)

Demo one project, show XML under the hood (Shelley Godwin?) 15 minutes (Elisa--Digital Mitford: prosopography (simple TEI to connect a large project)

Intro to markup/XML/TEI 20 min? EM? DO: EM share slides
(June) Introduction to markup, text encoding and XML
https://docs.google.com/presentation/d/1jJlHw88t7QtDUemazv3OieSkfyfpoJgMENKn5ehQnP8/edit?usp=sharing (up to slide 28)
Oxygen hands on - 30 mins - *EM develops outline*
, make sure they all have Oxygen
download / circulate files (network and/or thumb drive)
hands-on documentation (MS)
markup poem, phrase level, some attributes
(in handout, list which element we are going to talk about, include abbreviation, deletion/unclear, sic/corr, speech, quotes, names/places
Guidelines: show them and walk them through a decision-making process (at least one): names ("inline" markup vs. personography detail)

10:45-12:30
recap from first session - questions, clarifications 15-20 mins
Textual Editing approaches: variety supported by the TEI

- critical editions (comparing/documenting versions)

genetic markup: transcription/editing: writing process
TEIHeader -

- Digital Mitford *Charles the First* header
- concentrate on the required parts

--------Solid through here------
Different parts of the TEI - markup for different disciplines, purposes (Chapter list) related schemas (customizations) such as Epidoc, MEI ODD project level customization

- Show example ODD customizations.

Software and how to work with your files:

- XPath in oXygen first (maybe just navigating the descendant axis, with a predicate)
- eXist-db quick demonstration of XPath over a collection of documents via http://newtfire.org:8338 (elisa's training web installation of eXist-db)
- Maybe commentary on how eXist-db can help with project management / web development: Explanation of eXist-db integration with Digital Mitford letters: http://digitalmitford.org/letters.html (5 - 7 minutes)

- Links to more information.