# Navigating and Processing Data from the TEI with XSLT

**Half-day Workshop**  
Knowing how to locate and explore data in your encoding can help to learn how to work with TEI and XML generally. This workshop is designed for people who have some experience with TEI and seek to learn how to work with XML markup for analysis and research. Participants will gain a working, practical knowledge of the query language XPath and the transformation language XSLT, and learn how these can help to reduce reliance on software, packages and plugins that may become obsolete without warning. Further, XSLT's functional programming can serve as a way of articulating research questions around a document data model expressed in XML.

The emphasis of our workshop is “pull-processing”: that is, extracting data and metadata from markup documents for analysis, as opposed to providing the reading view of a digital scholarly edition. Markup in documents supplies structures and contexts that are especially useful for processing data, beyond what we can do with so-called "plain text". We will demonstrate some basic XPath navigation and calculation functions, and then show how XPath is applied in XSLT templates to address specific nodes that hold data of interest for visualization.

We will process TEI documents composed in Spanish and in languages represented by our workshop members' projects, to show that the code we write is transferable to multiple projects across language and cultural borders. Workshop instructors will collaborate and seek advice from the conference organizers on preparing Spanish-language source materials and documentation to establish an international foundation for this workshop.

Participants will learn how to "pull" data from TEI and output text formats required for simple online tools, where the structure of the output data is transferable to many different online calculation programs and amenable to statistical processing. During the workshop we will produce some simple structured documents for storing, sharing, and visualizing data: HTML lists and tables as well as plain text tabulated data (CSV or TSV files), and (if we have time) simple SVG bar or line graphs.

We hope to process some participant-supplied XML before, during, and after the workshop. We will carefully document the XSLT that we supply during the workshop to assist participants with revising and adapting the code to their own projects.

## Outline

* Review and refresh understanding of XML tree structures
* Orientation to XPath
* Teach basic XSLT to produce simple outputs ready for analysis and visualization

## Room/Materials Required

* Instructors need: projector that can connect to laptop, network access helpful!
* Participants should bring laptop computers if possible.
* If classroom with computers is available: provide guest login access to computers and install oXygen XML Editor.
* Instructors can provide complementary 90- or 120-day licenses for the oXygen XML Editor.
* Limit to 25 participants so instructors can connect with everyone.

## Workshop instructors

Elisa Beshero-Bondar, PhD

Program Chair of Digital Media, Arts, and Technology | Professor of Digital Humanities | Director of the Digital Humanities Lab at Penn State Erie, The Behrend College

An active member of the [Text Encoding Initiative (TEI)](http://www.tei-c.org/index.xml), Dr. Beshero-Bondar serves as an elected member and now chair of the [TEI Technical Council](http://www.tei-c.org/Activities/Council/), an eleven-member international committee that supervises amendments to the TEI Guidelines. She has been teaching humanities in web-savvy ways since the 1990s, and she began teaching markup languages and XML stack processing almost as soon as she began learning them in the 2010s. Before moving to direct the [DIGIT program](https://digit-psb.github.io/DIGIT/) at Penn State Erie, she directed Pitt-Greensburg's [Center for the Digital Text](http://www.greensburg.pitt.edu/digital-humanities/center-digital-text). She has led TEI data modeling of the [Frankenstein Variorum](https://frankensteinvariorum.org/) project, the [Digital Mitford Project](http://digitalmitford.org/) and other digital research projects involving TEI XML to build editions and prepare structured analyses of variants and collocations in texts. Find her on GitHub at <https://github.com/ebeshero> and on her development site named for her pet firebelly newts at <https://newtfire.org>.

Dr. Martina Scholger

Centre for Information Modelling - Austrian Centre for Digital Humanities, University of Graz

Martina Scholger has a PhD in Digital Humanities and holds a Senior Scientist position at the Centre for Information Modelling – Austrian Centre for Digital Humanities at the University of Graz. Her main research field is digital scholarly editing, the application of digital methods and semantic technologies to humanities’ source material, and text mining. 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 an elected member and past chair of the TEI Technical Council since 2016, and a member of the Institute for Documentology and Scholarly Editing since 2014. She has been teaching at a number of Summer Schools and workshops 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).