

Modelling humanities data with TEI-XML

Scholarly editing and manuscript cataloguing in the digital age

Dr Katarzyna Anna Kapitan (katarzyna.kapitan@chartes.psl.eu)

This course teaches students how to encode and model humanities data using TEI-XML, with a focus on creating digital scholarly editions and cataloguing manuscripts. Students will learn to model complex textual structures and metadata, enabling the digital analysis of gathered materials. Through practical exercises, participants will learn to apply the Text Encoding Initiative (TEI) guidelines to their own projects, as well as how to customise and document their encoding to fit their research questions. By the end of the course, students will be able to produce customised digital editions of texts and catalogue entries for manuscripts.

Keywords: data modelling, mark-up languages, digital scholarly editing, digital manuscript cataloguing, Extensible Markup Language (XML), Text Encoding Initiative (TEI), document design, schemas and customisation, DTD, ODD.

Pre-course preps:

This course will be taught in English, hence we strongly recommend that this summer you familiarise yourself with the TEI-XML-related terminology in English by reading the following publications:

- What is the Text Encoding Initiative?, available here <https://books.openedition.org/oep/426?lang=en> (you can read it side by side with its French translation)

If you have time, you can also give a look to the following:

- A gentle introduction to XML, available here: <https://tei-c.org/release/doc/tei-p5-doc/en/html/SG.html>
- Using the TEI, available here: <https://tei-c.org/release/doc/tei-p5-doc/en/html/USE.html>
- Critical apparatus, available here: <https://tei-c.org/release/doc/tei-p5-doc/en/html/TC.html>
- Manuscript description, available here: <https://tei-c.org/release/doc/tei-p5-doc/en/html/MS.html>
- *XML in a nutshell*, 2nd edition, available here <https://docstore.mik.ua/orelly/xml/xmlnut/index.htm> (especially chapters 1, 2, 3, 6, and 15)

Time: Wednesdays, 15:00-17:00

Contents:

Session 1	Markup Languages and Text Encoding
Session 2	Extensible Markup Language (XML) and Document Type Definition (DTD)
Session 3	Text Encoding Initiative (TEI) Guidelines
Session 4	Describing Primary Sources
Session 5	Transcribing Primary Sources
Session 6	Editing Primary Sources
Session 7	Annotating Primary Sources
Session 8	XPath
Session 9	Customisation and Documentation
Session 10	Customisation and Validation

Assessment:

- Attendance and Participation (20%)
- Encoding Project Portfolio:
 - Assignment 1 (due 20 October): 15%
 - Assignment 2 (due 17 November): 25%
 - Assignment 3 (due 22 December): 40%

Primary Readings

Week 1: Mark-up Languages and Text Encoding

- Hawkins, Kevin S., 'Introduction to XML for Text', <http://www.ultraslavonic.info/intro-to-xml/>.
- Birnbaum, David J., 'What is XML and why should humanists care? An even gentler introduction to XML' (2021) <http://dh.obdurodon.org/what-is-xml.xhtml>.

Week 2: Extensible Markup Language (XML) and Document Type Definition (DTD)

- Harold, Elliotte Rusty and Means, W. Scott, 'Chapter 3. Document Type Definitions (DTDs)'. In: *XML in a Nutshell*. (2022), <https://docstore.mik.ua/oreilly/xml/xmlnut/index.htm>.

Week 3: Text Encoding Initiative (TEI) Guidelines

- Burnard, Lou. 'The structural organization of a TEI document'. In: *What Is the Text Encoding Initiative?* (2014), <https://doi.org/10.4000/books.oep.426>.
- TEI Consortium, 'The TEI Header'. In: *TEI: Guidelines for Electronic Text Encoding and Interchange*, <https://tei-c.org/release/doc/tei-p5-doc/en/html/HD.html>.

Week 4: Describing Primary Sources

- TEI Consortium, 'Manuscript Description'. In: *TEI: Guidelines for Electronic Text Encoding and Interchange*, <https://tei-c.org/release/doc/tei-p5-doc/en/html/MS.html>

Week 5: Transcribing Primary Sources

- TEI Consortium, 'Default Text Structure'. In: *TEI: Guidelines for Electronic Text Encoding and Interchange*, <https://tei-c.org/release/doc/tei-p5-doc/en/html/DS.html>

Week 6: Editing Primary Sources

- TEI Consortium, 'Critical Apparatus' In: *TEI: Guidelines for Electronic Text Encoding and Interchange*, <https://tei-c.org/release/doc/tei-p5-doc/en/html/TC.html>

Week 7: Annotating Primary Sources

- TEI Consortium, 'Names, Dates, People, and Places'. In: *TEI: Guidelines for Electronic Text Encoding and Interchange*, <https://tei-c.org/release/doc/tei-p5-doc/en/html/ND.html>.

Week 8: XPath

- Harold, Elliotte Rusty and Means, W. Scott, 'Chapter 9. XPath'. In: *XML in a Nutshell*. (2022), <https://docstore.mik.ua/oreilly/xml/xmlnut/index.htm>.

Week 9: Customisation and Documentation

- TEI Consortium, 'Using the TEI'. In: *TEI: Guidelines for Electronic Text Encoding and Interchange*, <https://tei-c.org/release/doc/tei-p5-doc/en/html/USE.html>

Week 10: Customisation and Validation

- Beshero-Bondar Elisa E., 'Guide to Schema Writing with Schematron' (2020), <https://newtfire.org/courses/tutorials/explainSchematron.html>

Bibliography

- Beshero-Bondar Elisa E., 'Guide to Schema Writing with Schematron' (2020), <https://newtfire.org/courses/tutorials/explainSchematron.html>.
- Birnbaum, David J., Beshero-Bondar, Elisa E., Sperberg-McQueen, C. M., 'Flattening and unflattening XML markup: a Zen garden of XSLT and other tools' (2018), <https://www.balisage.net/Proceedings/vol21/html/Birnbaum01/BalisageVol21-Birnbaum01.html>
- Birnbaum, David J., 'What is XML and why should humanists care? An even gentler introduction to XML' (2021) <http://dh.obdurodon.org/what-is-xml.xhtml>.
- Burnard, Lou. *What Is the Text Encoding Initiative?* (2014).
- Bureau, Bruno, 'Quelques réflexions sur la notion de littérarité à partir de l'édition numérique de commentateurs anciens', *Interférences. Ars scribendi* (2012), <http://journals.openedition.org/interferences/186>.
- Bauman, Syd. 'Freedom to Constrain: where does attribute constraint come from, mommy?' In *Proceedings of Balisage: The Markup Conference 2008*. Balisage Series on Markup Technologies, vol. 1 (2008), <https://www.balisage.net/Proceedings/vol1/html/Bauman01/BalisageVol1-Bauman01.html>
- Cummings, James, 'The Present and Future of Encoding Text(s)'. In *The Bloomsbury Handbook to the Digital Humanities* (2022), https://eprints.ncl.ac.uk/file_store/production/288895/E8D00363-5822-40A1-B024-0306F3DA7360.pdf.
- Dodds, Leigh, *Schematron: validating XML using XSLT* (2001), http://ldodds.com/papers/schematron_xsltuk.html.
- Driscoll, Matthew, 'Levels of transcription', <https://tei-c.org/Vault/ETE/Preview/driscoll.html>.
- Driscoll, Matthew, 'Marking-up abbreviations in Old Norse-Icelandic', <http://www.driscoll.dk/docs/Abbreviations.pdf>
- Dufournaud, Nicole et al., 'Manuel d'encodage XML-TEI - édition numérique de manuscrits baroques' (2012), <https://hal.science/hal-00718043/document>.
- Duval, Frédéric, 'Pour des éditions numériques critiques. L'exemple des textes français', *Médiévales* (2017), <https://journals.openedition.org/medieval/8165>.
- Freeman, James and Holford, Matthew, et al., *TEI Manuscript Description ODD Customisation* (2023), <https://github.com/msDesc>.
- Harold, Elliotte Rusty and Means, W. Scott, 'Chapter 9. XPath'. In: *XML in a Nutshell*. (2022), <https://docstore.mik.ua/oreilly/xml/xmlnut/index.htm>.
- Haugen, Odd Einarm et al., *The Menota handbook: Guidelines for the electronic encoding of Medieval Nordic primary sources* (2019), <http://www.menota.org/handbook.xml>.
- Hawkins, Kevin S., 'Introduction to XML for Text', <http://www.ultraslavonic.info/intro-to-xml/>.
- Heiden, Serge et al., 'Manuel d'encodage XML-TEI des textes de la Base de Français Médiéval' (2010), http://bfm.ens-lyon.fr/IMG/pdf/Manuel_Encodage_TEI.pdf.
- Honkapohja, Alpo, 'Manuscript abbreviations in Latin and English: History, typologies and how to tackle them in encoding', <https://varieng.helsinki.fi/series/volumes/14/honkapohja/>.
- Kay, Michael R., *XSLT 2.0 and XPath 2.0 Programmer's Reference* (2011).
- Lavignasse, Sophie, 'La DTD et son langage XML. Une application pour la lexicographie contemporaine', *Éla. Études de linguistique appliquée* (2005), <https://www.cairn.info/revue-ela-2005-1-page-73.htm>.
- Ohge, Christopher, *Publishing scholarly editions* (2021).
- Palluault, Florent, 'Informatiser des descriptions complexes: l'utilisation de l'EAD et de la TEI pour les manuscrits et les livres anciens en France', *IFLA* (2012), <https://www.ifla.org/past-wlic/2012/212-palluault-fr.pdf>.
- Pierazzo, Elena, *Digital scholarly editing: theories, models and methods* (2015), <https://hal.univ-grenoble-alpes.fr/hal-01182162/document>.

- Rahtz, Sebastian and Burnard, Lou, 'Reviewing the TEI ODD system'. In *Proceedings of the 2013 ACM symposium on Document engineering* (2013), <https://doi.org/10.1145/2494266.2494321>.
- Robinson, Peter, 'Where We Are with Electronic Scholarly Editions, and Where We Want to Be', *Jahrbuch für Computerphilologie* (2003), <http://computerphilologie.digital-humanities.de/jg03/robinson.html>.
- Robinson, Peter and Solopova, Elizabeth, 'Guidelines for Transcription of the Manuscripts of the Wife of Bath's Prologue'. In *Occasional Papers* (1993), <http://canterburytalesproject.com/pubs/op1-transguide.pdf>.
- Siegel, Erik, *Schematron: A Language for Validating XML* (2022).

Digital projects

Start by checking out what's in the TEI inventory of projects, <https://tei-c.org/activities/projects/>

- Aratea Digital, <https://ivanadob.github.io/aratea-data/index.html>
- Beta maṣāḥəft: Manuscripts of Ethiopia and Eritrea, <https://betamasaheft.eu/index.html#manuscripts>
- Bodleian First Folio, <https://firstfolio.bodleian.ox.ac.uk/text/>
- Handrit, <https://handrit.is>
- Les éditions électroniques de l'École nationale des chartes, <http://elec.enc.sorbonne.fr>
- Medieval Nordic Text Archive, <https://clarino.uib.no/menota/catalogue/menota>
- Newton Project, <https://www.newtonproject.ox.ac.uk>
- Online Froissart, <https://www.dhi.ac.uk/onlinefroissart/>
- Perseus Digital Library, <http://www.perseus.tufts.edu/hopper/>
- Stories for all time, <https://fasnl.net>
- Taylor Editions, <https://editions.mml.ox.ac.uk/editions/>
- The Shelley-Godwin Archive, <http://shelleygodwinarchive.org/about/>

Self-study materials

- Albouy, Segolene, *Cours XML-TEI, Technologies Numériques Appliquées à l'Histoire* (2024), https://github.com/Segolene-Albouy/XML-TEI_M2TNAH.
- Bauman, Syd, Flanders, Julia, et al., *Women Writers Online TEI Tutorials* (2007–), https://www.wwp.northeastern.edu/outreach/resources/tutorial_all.html.
- Beshero-Bondar, Elisa E. and Birnbaum David J., *Code the X-files using the XML family of languages* (2024), https://ebeshero.github.io/UpTransformation/dhsi-XPath_CourseDescription.html.
- Burghart, Marjorie, *TEI Critical Apparatus Toolbox* (2016), <http://teicat.humanum.fr/credits.php>.
- Burghart, Marjorie, Pierazzo, Elena et al., *Digital Scholarly Editions: Manuscripts, Texts and TEI Encoding* (n.d.), <https://teach.dariah.eu/course/view.php?id=32>.
- Carton, Olivier, *L'essentiel de XML* (2015), <https://www.irif.fr/~carton/Enseignement/XML/Cours/index.html>.
- Chateau, Emmanuel, *L'encodage des textes*, (2012), <http://www.desgodets.net/edition-des-cours/model>.
- Pinche, Ariane, *Cours XML TEI - Master TNAH* (2018), https://github.com/ArianePinche/coursTNAH_XML-TEI.
- TEI Consortium, *TEI: Guidelines for Electronic Text Encoding and Interchange* (Version 4.8.0. Last updated on 2nd September 2024,) <https://tei-c.org/release/doc/tei-p5-doc/en/html/index.html>.

- TEI Consortium, 'Getting Started with P5 ODDs', <https://web.archive.org/web/20230421200024/https://tei-c.org/guidelines/customization/getting-started-with-p5-odds/>.
- TEI Consortium, 'Customizing the TEI with Roma', <https://web.archive.org/web/20230201010911/https://tei-c.org/guidelines/customization/customizing-the-tei-with-roma/>.
- Van Den Branden, Ron, Terras, Melissa and Vanhoutte Edward, *TEI by Example* (n.d.), <http://www.teibyexample.org>.
- XPath Tester, <https://extendsclass.com/xpath-tester.html>.