

The TEI toolkit

Lou Burnard Consulting

Sept 2018



What can you do with a TEI XML file?

- The TEI itself doesn't provide you with a toolbox, nor a single do-everything package
- The TEI claims to be an application-independent standard: it cannot therefore at the same time propose applications
- The **Guidelines** are designed to give concrete expression of an abstract model of the objects – mostly textual – which are of scientific interest to the Humanities community...
- .. but only the members of that community can determine how to process such objects, and hence how to build the tools to make use of them.

What sort of TEI-XML tools do we need?

What might you want to do with a TEI XML document?

- create, modify, validate ...
- transform, display, visualise ...
- search, analyse, mash-up ...
- store, preserve, archive, catalogue ...

Word to TEI

With oXygen, for example

- Use the usual File Open (ctrl-o) dialog to select any docx file
- A window labelled Archive Browser opens to the left of the main screen. It shows the file structure of the docx archive.
- Click the blue key next to the folder called word to see the contents of this folder
- Select the file called document.xml and double click to open it (may take a few moments if the file is large)
- With the document.xml file open in your main editing window, select Transformation -> Configure Transformation Scenario(s) from the Document menu. Or type CTRL - SHIFT - C. Or click the little spanner icon
- For the default DOCX to TEI conversion, check the little box next to DOCX TEI P5 and press the Apply Associated button

This can also be done at the command line, or online using oxGarage



Creation and modification of TEI XML documents

- Editors like oXygen are not unique !
- See http://wiki.tei-c.org/index.php/Category:Editing_tools for a range of others, some of them free – or just ask Google
- Documents created with ordinary office tools (Word, Open Office) can be automatically converted to TEI XML
- Highly structured documents (metadata for example) can be captured by a form and output directly in TEI XML

You are not alone

The TEI Guidelines are written in XML, like every other TEI application

- so any XML-aware software can be used to process them
- but such software needs to be customized!

For its own needs, the TEI produced and now maintains a suite of XSLT stylesheets which supports :

- generation and documentation of customised TEI schemas ("ODD")
- visualisation of arbitrary TEI documents, using commonly used formats (XHTML, PDF, Word, Open Office, ePUB...)
- a generic transformation architecture, supporting TEI and other formats

Transformation and visualisation of TEI-XML documents

The TEI Stylesheets library originally developed by Sebastian Rahtz and now maintained by the TEI:

- packaged by oXygen as a Framework
- free download from gitHub
(<https://github.com/TEIC/Stylesheets>)
- component of a TEI-hosted web service EGE
- integrated with many applications developed independently of the TEI
- maintained and developed by the TEI (since the TEI uses it internally)

Transformations already provided

	from TEI	to TEI
Office document formats		
OOXML (docx), ODF (odt), Docbook	*	*
LaTeX, PDF	*	
ePub, XSLFO	*	
Web formats		
HTML5, JSON, RDF	*	
Wordpress		*
Markdown	*	*
Schema languages (via ODD)		
XML DTD, RELAXNG, Schematron, W3C Schema	*	
Other XML formats		
TEI P4		*
NLM	*	
Verbatim XML	*	*
Non-XML formats		
Cocoa	*	
Plain text	*	

Customizing the TEI Stylesheets

The Stylesheets are designed to be customized...

- Profile** you can set up a named bunch of transformations and store it as a *framework* within oXygen (several examples come with the product)
- CSS** you can use your own CSS stylesheet/s to manage details of how the output will be displayed, on screen or in print
- Stylebear** you can use the styleBear application to simplify generation of a customized stylesheet
- LaTeX** LaTeX users can generate high quality PDF output (requires knowledge of LaTeX)

How should you publish XML-TEI resources?

The policy of least effort...

- Here are our XML-TEI files. Go figure.
<http://www.cnrtl.fr/corpus/estrepublikain/>
- The Archimedes palimpsest
<http://archimedespalimpsest.net>
- Oxford Text Archive (<http://ota.ox.ac.uk/>)
- handling multiple scripts: Samyukta Agama (<http://buddhistinformatics.ddbc.edu.tw/BZA/>)

Digital publishing systems

For the management, storage, search, and display of digital TEI editions there is now a variety of software solutions :

- Plug-ins for common CMS (Drupal, Zotero, Omeka)
- TEI specific systems (Kiln, TEI Boilerplate, CETEICEAN)
- General purpose document management systems (xsf, TEI Publisher)

For the most part, these are systems aimed at web developers, not end-users

XTF : a digital library creator (for example)

- Extensible Text Framework or XTF
(<http://www.cdlib.org/inside/projects/xtf/>)
is a collection of server-side scripts from the California Digital Library
- If you are already running Apache and Tomcat, and have access to a website, you can set up a default xtf application for TEI files in 5 minutes
- (more or less)
- If not, the infrastructural overhead may seem prohibitive...

Some less-demanding tools

- TEI Boilerplate
- CETElcean
- TEI Publisher

These require knowledge of how to upload pages to a website. That is all.

TEI Publisher (for example)

- visit `http://http://tei-publisher.com` and click the try it button
- To upload a file, click "login" button (default user: "tei", password: "simple")
- scroll down to "Upload" dialogue (bottom right)
- Select a TEI XML file from the Work directory (try `suprasliensis-tei.xml`)
- When the file is uploaded, click the goto link to see it displayed in various formats

To control the way your file is formatted, you need to supply an ODD containing processing instructions : so you need only TEI knowledge

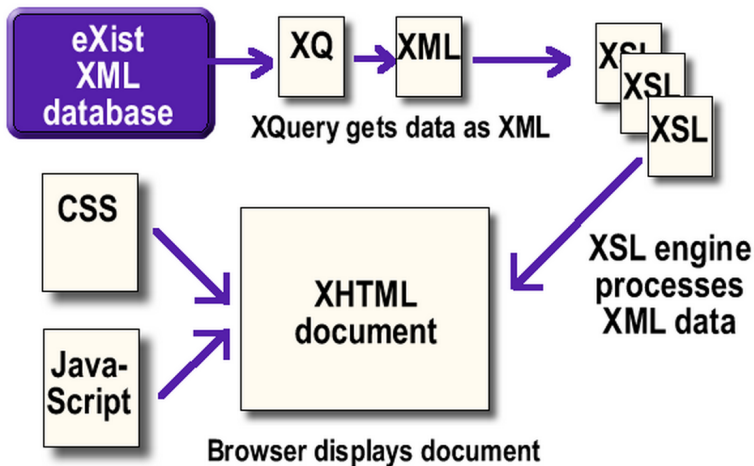
Bottom line: use the markup!

To take full advantage of the XML markup (for example, to search and analyse your documents in terms of their markup) you are usually better off investing in a generic XML database system, such as

- baseX <http://basex.org>
- eXist <http://exist-db.org>

The solution par excellence for any project managing serious amounts of TEI XML data

A typical software architecture



Examples...

- Colonial Despatches:
<http://bcgenesis.uvic.ca/docsByDate.htm>
- Carl Maria von Weber
Archive <http://weber-gesamtausgabe.de/en/A002068/Correspondence>
- Ancient Wisdoms : kiln <http://www.ancientwisdoms.ac.uk/method/software-install/>
- Shelley-Godwin archive : shared canvas viewer
<http://shelleygodwinarchive.org/about>
- Letters and mss of 19th c Berlin <http://tei.ibi.hu-berlin.de/berliner-intellektuelle/manuscript?Sandmann+en#5>
- Bibliotheque Virtuelle des Humanistes : philologica
<http://www.bvh.univ-tours.fr/Epistemon/philologic.asp>

Research tools

Specialised tools developed for or by particular research communities are also increasingly using TEI as a base format or as an input format:

- 'textometrie' software for lexical statistics
- many packages for critical editions
- analysis and representation of spoken language
- linguistic analysis in general
- prosopographic information extracted from archival sources

More examples ...

<http://www.tei-c.org/Activities/Projects>

Conclusions

- even the most minimal of approaches makes it possible for you to share your personal analyses of a document
- the more elaborate your markup, the more you can do with it
- there is however a common core of techniques and facilities: no need to reinvent the wheel
- TEI XML empowers the data provider : it's up to you to decide how your materials are exposed and accessed