TEI by example project team meeting, KANTL, 2006/12/01

Present: Edward Vanhoutte (EV), Melissa Terras (MT)

Minutes: Ron Van den Branden (RVDB)

1. State of affairs

The meeting started with a presentation of following things done so far:

- Publication of initial project report [TBE-R001], 2006/06/09 and announcement of the project on the TEI mailing list. This generated one useful suggestion by Serge Heiden to help with French translation of the deliverables.
- The current XML tools list has been XML-ised (TEI P5) and is ready for update.
- Initial content, mirroring the current TEI by example website has been set up at the Centre for Computing in the Humanities (CHH), to be delivered by a server hosting a Cocoon web application + FOP XSL-FO processor and eXist native XML database.
- 2006/09/18 – 2006/09/20: RVDB and MT attended a TEI training workshop at Oxford University Computing Services.

So far, 10 of the budgeted 90 days were spent at the project. This leaves 80 more days, or 32 weeks of 2,5 days devoted to the project. Overall, the date for the Digital Humanities 2007 Conference is considered a good target for finalisation of the tutorial development phase.
Further meetings are planned, either 1 halfway the development period (2007/03), or 2 at 2007/02 and 2007/04.

2. Pending issues, identified in [TBE-R001]

2.1 TEI version

The suggestion in [TBE-R001] was agreed on, to use TEI P5 (version 0.5 at the moment of writing) and build in a revision round for updating the teaching materials at the end of the project where necessary. This will guarantee that the tutorials will conform to the current version of TEI P5.

2.2 XML editor

[TBE-R001] identified the need for a dedicated XML editing environment for reference purposes in the tutorials, and proposed the freely available Exchanger XML Lite editor. This proposal was accepted, with some refinement:

- RVDB relativised the need for an XML editor to be most prominent in the introductory teaching module (most examples will be just standalone XML fragments).
- EV proposed to develop a worked out tutorial for text encoding with the proposed software, and ask the members of the project's advisory committee to 'translate' this to the XML editor they use.

2.3 Creative Commons licence

[TBE-R001] identified the need for a licence arrangement that allows the dissemination of the (example fragments in the) tutorials for academic purposes. Following non-commercial Creative Commons (CC) licences were proposed:

- Attribution Non-commercial No Derivatives (by-nc-nd)
  allows unmodified redistribution under identical terms, for non-commercial purposes
- Attribution Non-commercial Share Alike (by-nc-sa)
  allows modified redistribution under identical terms, for non-commercial purposes
- Attribution Non-commercial (by-nc)
  allows modified redistribution under different (yet non-commercial) terms

Decision was made for the most restrictive, viz. Attribution, Non-commercial No Derivatives licence. MT pointed out the possibility of free institutional advice on copyright matters, if required.

2.4 TEI endorsement – future development

Communication with the TEI board revealed an interest from the TEI to be associated with the work of this project. Although TEI endorsement would be very valuable for both this project and the TEI itself, it is deemed too early to decide on the exact nature and degree of such endorsement. After completion of this project, the work done could be valorized by guaranteeing future sustainability and close integration with the TEI text encoding community, via e.g. the TEI Wiki. In a more substantive scenario, TEI-certified on line quizzes and tests accompanying the tutorials could form the basis for some form of 'TEI driving license'. This is deferred to a later stage, when the finished product will be submitted to the TEI board.

Related to this issue, MT brought up the issue of a future application with ALLC for maintenance and updating of the TEI by example tutorials after completion of the this project. The next deadline for such applications is November 2007.

2.4 TEI schema

Regarding the scope of the TEI schema to be used in the tutorial modules, two options remained:

- a monolithic, comprehensive ODD for all modules (cf. *tei_all.odd*)
- dedicated ODDs per module, containing a very specific subset

The second option (forcing a motivated selection of relevant elements) was considered most reasonable, provided that the status of the ODD file is made explicit in a disclaimer for each module. On a technical side, the choice for one of different expression formats of each module's TEI ODD file (DTD / W3C Schema / Relax NG Schema) to be offered for download was considered irrelevant. Instead of just one format, the whole constellation of an ODD file and its derived expression formats should be offered for download, with reference to module 8 (Customizing TEI, ODD, Roma).

3. Example fragments

3.1 Gathering of example fragments

A call for examples features high on the to-do-list, since real life examples from the text encoding community are vital for the development of the project. A number of issues were identified and discussed, which informed a proposed general call for examples, as attached in appendix A to these minutes. A first issue addressed the optimal approach to get what we want, viz. relevant TEI P5 encoded texts or text fragments for each of the module subjects. Considered the relative newness of TEI P5 at the time of writing, few projects can be expected to have TEI P5 encoded texts. Therefore, it would be more realistic to broaden the call for examples to TEI P4 XML texts, any XML encoded texts, or otherwise electronically available texts (plain text / Word / PDF) as well.

Although English would be the current language of focus (translation of the course materials and multilingual examples is a future action), explicating this in a call for examples was considered unnecessarily controversial. For the practical organisation of calls for examples, a two-tiered approach was decided on:

1. A general call for examples at the start of the development phase (as soon as possible).
2. Specific, targeted calls for examples before the start of each module. A period of two weeks before the start of each module's development was proposed for specific calls and editing of the collected examples, unless this would prove too tight.

Examples will be gathered by email, via teibyexample@kantl.be. Following partners were identified for contributing example fragments:

Call for examples	Who	Awareness Raising
1. Global call	Text encoding communities	Mailing lists: TEI-L[1], Humanist[2], Digital Medievalist[3], Digital Classicist[4] Blogs: Grand Text Auto[5]
	Dedicated XML repositories	Oxford Text Archive[6]
	Projects using TEI	Email of 135 separate projects listed at <http://www.tei-c.org/Applications/>
	Project advisory committee	Personal email
2. Specific calls	Project advisory committee	Personal email

To maximise responsiveness for the global call for examples, it must feature the stance regarding copyright, as well as the option to provide imperfect examples, and the fact that contributors will be credited in a general colophon. These points of attention have been taken into consideration for the proposed global call for examples, attached in the appendix A to these minutes.

3.2 Didactic approach

Depending on the quality and quantity of submitted examples, following options can be considered:

- 1 incrementally worked out example throughout a module
- several parallel examples per module

Although the latter option obviously requires many examples (which could be problematic with typically low responsiveness for calls for TEI encoded materials), EV pointed out that the former option holds the danger of monopoly of 1 period, genre, author and thus practice of one text encoding community in the examples. Ways to avoid this could be:

- the ability to present a rotating pool of examples
- construction of fictitious composite examples from different periods, authors etc.

Both options have their problems: the first one would require nearly sterile descriptive prose in the tutorials, that would allow the insertion of random example fragments. The second option would

1 TEI-L: <http://listserv.brown.edu/archives/tei-l.html>
2 Humanist: <http://www.princeton.edu/~mccarty/humanist/>
3 Digital Medievalist: <http://www.digitalmedievalist.org/mailing.cfm>
4 Digital Classicist: <http://www.jiscmail.ac.uk/lists/digitalclassicist.html>
5 Grand Text Auto: <http://grandtextauto.gatech.edu/>
6 Oxford Text Archive: <http://ota.ahds.ac.uk/>

probably quickly alienate readers from the core didactic concepts the examples should convey.

MT proposed an alternative: the use of an incremental example throughout a module, concluded with a quick overview of other worked out examples at the end.

Another issue is the degree of interactivity in the tutorials. Ideally, the tutorials could be interspersed with interactive tests that would allow the readers to get relevant feedback for their encoded text fragments. The feasibility of such a set-up is yet to be investigated, both from a technical point of view (full XML validation vs. shallow string comparison) and a didactic point of view (semantic XML validation is a problematic notion, and how didactic are generic XML parser error messages?).

4. Development

In the meeting, a revised project development plan was proposed, aiming at completion near the Digital Humanities 2007 conference, from 4 to 7 June (28 weeks). The GANTT chart is added in appendix B to these minutes, comments are discussed below.

4.1 Identify module goals

Prior to the actual work, it's important to get a clear view on each module's goals and the overall approach. The presentation of an initial proposal for module goals (indicating desirable topics and elements to be treated) evoked a fairly broad discussion of general issues. An important issue in this debate was the approach towards overlap of didactic contents. Since the TEI is organised in a modular way, lots of TEI elements are shared in the separate topics addressed by the tutorial modules. EV suggested to focus on mandatory tags, MT proposed to select 10 to 12 additional optional ones to be addressed in the modules. This still leaves the problem that every TEI document must conform to a common minimal structure (mandatory elements, possibly containing conditionally mandatory substructures), and can contain lots of common optional elements. Two possible approaches are to be investigated:

- include the common ground in each module, with dedicated examples
- separate out the common ground into a distinct module, to be referred to from the other modules

Since a detailed analysis of the initial proposal would have exceeded the available time, this is formally deferred to January 2007, when EV and RVDB will devote their time to the start of the actual module development. In order to provide a start for informal reflection, the initial proposal is included here. Each module is followed by

- a list of proposed concepts and tags
- a full list of elements[7] defined in the modules identified in [TBE-R001]

7 Elements marked with a yellow background are mandatory in all TEI modules. Elements with a green background indicate suggested optional elements.

	Module	Concepts	Tags
1	Introduction	- Procedural vs descriptive markup	
		- XML – well-formed XML – valid XML (DTD / Schema ~> ODD)	
		- General TEI document structure (different genres)	<TEI>, <teiHeader>, <text>, <body>, <front>, <back>, , <group>, <head>
		- Global attributes	@xml:id, @xml:lang, @rend, @n
		- Low-level descriptive tags	<name>, <date>, <figure>, <ref>
		- TEI modular architecture	
	ch. 7	TEI text body group div div0 div1 div2 div3 div4 div5 div6 div7 trailer byline dateline argument epigraph opener closer salute signed titlePage docTitle titlePart docAuthor imprimatur docEdition docImprint docDate front back	
	ch. 6	p foreign emph hi distinct q quote cit mentioned soCalled altIdent desc equiv gloss term sic corr choice reg orig gap add del unclear name rs address addrLine street postCode postBox num measure date time abbr expan ptr ref list item label head headLabel headItem note index divGen graphic binaryObject milestone pb lb cb bibl biblItem biblStruct biblFull listBibl analytic monogr series author editor respStmt resp title meeting imprint publisher biblScope pubPlace l lg sp speaker stage teiCorpus	

	Module	Concepts	Tags
2	TEI header	- General components of standard TEI header - Minimal recommended header	<teiHeader>, <fileDesc>, <encodingDesc>, <profileDesc>, <revisionDesc> + offspring
	teiHeader fileDesc titleStmt sponsor funder principal editionStmt edition extent publicationStmt distributor authority idno availability seriesStmt notesStmt sourceDesc scriptStmt recordingStmt recording equipment broadcast encodingDesc projectDesc samplingDecl editorialDecl correction normalization quotation hyphenation segmentation stdVals interpretation tagsDecl tagUsage namespace rendition refsDecl cRefPattern state classDecl taxonomy category catDesc fsdDecl metDecl metSym variantEncoding profileDesc creation langUsage language textClass keywords classCode catRef revisionDesc change		

	Module	Concepts	Tags
3	Prose	- General structure of prose	, <head>, <p>, <q>
		- Low-level descriptive tags	<name>, <date>, <address>, <title>, <ref>
		- Other specific elements	<figure>, <table>
	ch. 6	p foreign emph hi distinct q quote cit mentioned soCalled altIdent desc equiv gloss term sic corr choice reg orig gap add del unclear name rs address addrLine street postCode postBox num measure date time abbr expan ptr ref list item label head headLabel headItem note index divGen graphic binaryObject milestone pb lb cb bibl biblItem biblStruct biblFull listBibl analytic monogr series author editor respStmt resp title meeting imprint publisher biblScope pubPlace l lg sp speaker stage teiCorpus	
	ch. 7	TEI text body group div div0 div1 div2 div3 div4 div5 div6 div7 trailer byline dateline argument epigraph opener closer salute signed titlePage docTitle titlePart docAuthor imprimatur docEdition docImprint docDate front back	
	ch. 14	link linkGrp ab anchor seg when timeline join joinGrp alt altGrp	
	ch. 20	persName surname forename genName nameLink addName roleName placeName bloc country region district settlement offset distance geogName geog orgName orgTitle orgType orgDivn affiliation age birth death education faith floruit langKnowledge langKnown listPerson nationality occupation particLinks persEvent persState persTrait person personGrp relation residence sex socecStatus day week month year occasion second minute hour	

	Module	Concepts	Tags
4	Poetry	- General structure of poems	, <head>, <lg>, <l>
		- Specific poetry elements	<caesura>, <rhyme>, @rhyme, @enjamb
	ch. 6	p foreign emph hi distinct q quote cit mentioned soCalled altIdent desc equiv gloss term sic corr choice reg orig gap add del unclear name rs address addrLine street postCode postBox num measure date time abbr expan ptr ref list item label head headLabel headItem note index divGen graphic binaryObject milestone pb lb cb bibl biblItem biblStruct biblFull listBibl analytic monogr series author editor respStmt resp title meeting imprint publisher biblScope pubPlace l lg sp speaker stage teiCorpus	
	ch. 7	TEI text body group div div0 div1 div2 div3 div4 div5 div6 div7 trailer byline dateline argument epigraph opener closer salute signed titlePage docTitle titlePart docAuthor imprimatur docEdition docImprint docDate front back	
	ch. 8	caesura rhyme	
	ch. 14	link linkGrp ab anchor seg when timeline join joinGrp alt altGrp	
	ch. 20	persName surname forename genName nameLink addName roleName placeName bloc country region district settlement offset distance geogName geog orgName orgTitle orgType orgDivn affiliation age birth death education faith floruit langKnowledge langKnown listPerson nationality occupation	

		particLinks persEvent persState persTrait person personGrp relation residence sex socecStatus day week month year occasion second minute hour

	Module	Concepts	Tags
5	Drama	- General structure of performance texts	, <head>, <sp>
		- Specific performance text elements	<speaker>, <l>, <p>, <stage>
	ch. 6	p foreign emph hi distinct q quote cit mentioned soCalled altIdent desc equiv gloss term sic corr choice reg orig gap add del unclear name rs address addrLine street postCode postBox num measure date time abbr expan ptr ref list item label head headLabel headItem note index divGen graphic binaryObject milestone pb lb cb bibl biblItem biblStruct biblFull listBibl analytic monogr series author editor respStmt resp title meeting imprint publisher biblScope pubPlace l lg sp speaker stage teiCorpus	
	ch. 7	TEI text body group div div0 div1 div2 div3 div4 div5 div6 div7 trailer byline dateline argument epigraph opener closer salute signed titlePage docTitle titlePart docAuthor imprimatur docEdition docImprint docDate front back	
	ch. 9	set prologue epilogue performance castList castGroup castItem role roleDesc actor move view camera sound caption tech	
	ch. 14	link linkGrp ab anchor seg when timeline join joinGrp alt altGrp	
	ch. 20	persName surname forename genName nameLink addName roleName placeName bloc country region district settlement offset distance geogName geog orgName orgTitle orgType orgDivn affiliation age birth death education faith floruit langKnowledge langKnown listPerson nationality occupation particLinks persEvent persState persTrait person personGrp relation residence sex socecStatus day week month year occasion second minute hour	

	Module	Concepts	Tags
6	Manuscript Transcription	- Manuscript description header	<msDescription> + offspring
		- Manuscript description phrase level elements	<catchwords>, <dimensions>, <heraldry>, <locus>, <material>, <watermark>, <origDate>, <origPlace>, <secFol>, <signatures>
		- Primary source markup tags	<add>, <del>, <unclear>, <supplied>, <choice>, <abbr>, <expan>, <orig>, <reg>
	ch. 6	p foreign emph hi distinct q quote cit mentioned soCalled altIdent desc equiv gloss term sic corr choice reg orig gap add del unclear name rs address addrLine street postCode postBox num measure date time abbr expan ptr ref list item label head headLabel headItem note index divGen graphic binaryObject milestone pb lb cb bibl biblItem biblStruct biblFull listBibl analytic monogr	

		series author editor respStmt resp title meeting imprint publisher biblScope pubPlace l lg sp speaker stage teiCorpus
	ch. 7	TEI text body group div div0 div1 div2 div3 div4 div5 div6 div7 trailer byline dateline argument epigraph opener closer salute signed titlePage docTitle titlePart docAuthor imprimatur docEdition docImprint docDate front back
	ch. 13	msDescription catchwords dimensions height depth width heraldry locus material origDate origPlace secFol signatures watermark msIdentifier institution repository collection altIdentifier msName colophon explicit filiation finalRubric incipit msContents msItem msItemStruct rubric summary textLang physDesc objectDesc supportDesc support collation foliation condition layoutDesc layout handDesc handNote musicNotation decoDesc decoNote additions bindingDesc binding sealDesc seal accMat history origin provenance acquisition additional adminInfo recordHist source custodialHist custEvent surrogates msPart
	ch. 14	link linkGrp ab anchor seg when timeline join joinGrp alt altGrp
	ch. 17	certainty respons
	ch. 18	addSpan delSpan restore supplied hand handShift handList damage space fw
	ch. 20	persName surname forename genName nameLink addName roleName placeName bloc country region district settlement offset distance geogName geog orgName orgTitle orgType orgDivn affiliation age birth death education faith floruit langKnowledge langKnown listPerson nationality occupation particLinks persEvent persState persTrait person personGrp relation residence sex socecStatus day week month year occasion second minute hour

NOTES:

- how introductory is this topic (I don't know much about this transcription praxis...)
- VERY comprehensive: <msDescription>: possibly very complex header structure (although it can get as simple as it needs)
  msDescription.content =
  msIdentifier,
  head*,
  (
  model.pLike+
  | (msContents?, physDesc?, history?, additional?, msPart*)
)

	Module	Concepts	Tags
7	Scholarly Editing	- Apparatus + parallel segmentation (OR all methods?)	<app>, <rdg>, <lem>, <witList>, <witness>
		- Primary source markup tags	<add>, <del>, <unclear>, <supplied>, <choice>, <abbr>, <expan>, <orig>, <reg>
	ch. 6	p foreign emph hi distinct q quote cit mentioned soCalled altIdent desc equiv gloss term sic corr choice reg orig gap add del unclear name rs address addrLine street postCode postBox num measure date time abbr expan ptr ref list item	

		label head headLabel headItem note index divGen graphic binaryObject milestone pb lb cb bibl biblItem biblStruct biblFull listBibl analytic monogr series author editor respStmt resp title meeting imprint publisher biblScope pubPlace l lg sp speaker stage teiCorpus
	ch. 7	TEI text body group div div0 div1 div2 div3 div4 div5 div6 div7 trailer byline dateline argument epigraph opener closer salute signed titlePage docTitle titlePart docAuthor imprimatur docEdition docImprint docDate front back
	ch. 14	link linkGrp ab anchor seg when timeline join joinGrp alt altGrp
	ch. 17	certainty respons
	ch. 19	app lem rdg rdgGrp witDetail wit witList witness witStart witEnd lacunaStart lacunaEnd
	ch. 20	persName surname forename genName nameLink addName roleName placeName bloc country region district settlement offset distance geogName geog orgName orgTitle orgType orgDivn affiliation age birth death education faith floruit langKnowledge langKnown listPerson nationality occupation particLinks persEvent persState persTrait person personGrp relation residence sex socecStatus day week month year occasion second minute hour

	Module	Concepts	Tags
8	Customizing TEI, ODD, Roma	- Global organisation of TEI grammar (modules, elements, attributes) - Expression in ODD format	<schemaSpec>, <elementSpec>, <classSpec>, <macroSpecq>, <attList>, <attDef>, <desc>, <valList> <valItem>
		- Roma - TEI conformance	Roma tour (cf. Laurent Romary OUCS)
	ch. 7	TEI text body group div div0 div1 div2 div3 div4 div5 div6 div7 trailer byline dateline argument epigraph opener closer salute signed titlePage docTitle titlePart docAuthor imprimatur docEdition docImprint docDate front back	
	ch. 27	att code eg egXML gi ident tag val specList specDesc moduleRef moduleSpec schemaSpec specGrp specGrpRef stringVal elementSpec classSpec macroSpec remarks listRef exemplum classes memberOf content attList attDef attRef datatype defaultVal valDesc valItem valList	

NOTE: TEI conformance is a highly unstable / controversial issue at the time of writing!

4.2 + 4.3: Call for examples – examples processing

A general call for examples will be issued via the channels identified in section 3.1 of these minutes as soon as possible in December 2006. It is hoped that this will produce a number of usable example fragments, that can be analysed, edited and marked up in the first weeks of the development phase. Before the start of the development for each module, a specific call for examples will be targeted at individual persons, at least 2 weeks before the start of the module's development.

4.4 Tools list

The current listing of XML tools and resources has been XML-ised and is ready for revision and updating. This can be completed in 1 week.

4.5 - 4.12 Development + revision of modules

During the development phase, each module's content will be published internally for the project team. After completion of each module, 2 weeks are reserved for revision by the project team.

4.13 Project Report

It is estimated that the authoring of the project report will provide a good starting point for the poster submitted for the DH2007 conference.

4.14 XSLT development

Two issues need resolution:

- Is there an argument against using the TEI stylesheets?
- Prior to stylesheet development, views should be identified on the desired layout and functionality.

4.15 Publication

At the moment of writing, the CHH have configured a server environment (Cocoon) that allows dynamic delivery of XML contents via XSLT (Xalan), XSL-FO transformations (FOP), and XQuery (eXist). A dedicated mechanism for file transfer and publication will be worked out. Following publication cycle is proposed:

- During development, the modules will be published internally, with access rights for the project team. For each module, an immediate revision period of 2 weeks is scheduled..
- Final publication would consist of making the developed contents publicly accessible. stage: open up deliverables to wide public

A final question is whether an intermediary phase should be provided for feedback by the project advisory committee.

5. Action points

Following are the main immediate action points decided on in this meeting:

- RVDB: issue a general call for examples as soon as possible in December 2006. The call will be sent to the partners identified in section 3.1 of these minutes
- RVDB + EV: identify detailed module goals in the first weeks of January 2007

Appendix A: proposed call for examples

Subject: TEI by example: CALL FOR EXAMPLES

[apologies for cross-posting]

The Centre for Scholarly Editing and Document Studies (CTB) <http://www.kantl.be/ctb/> of the Royal Academy of Dutch Language and Literature, the Centre for Computing in the Humanities (CCH) <http://www.kcl.ac.uk/humanities/cch/> of King's College London, and the School for Library, Archive, and Information Studies (SLAIS) <http://www.slais.ucl.ac.uk/> of University College London, are involved in the joint project "TEI by Example".

Featuring freely available online tutorials walking individuals through the different stages in marking up a document in TEI (Text Encoding Initiative <http://www.tei-c.org>), these online tutorials will provide examples for users of all levels. Examples will be provided of different document types, with varying degrees in the granularity of markup, to provide a useful teaching and reference aid for those involved in the marking up of texts.

Eight tutorial modules will address a wide range of issues in text encoding with TEI:

1. Introduction to text encoding with TEI
2. The TEI header
3. Prose
4. Poetry
5. Drama
6. Manuscript Transcription
7. Scholarly Editing
8. Customizing TEI, ODD, Roma

To build as much as possible on available sources of existing practice in the field and to be able to present a broad view on the wide variety of encoding practices, we warmly welcome you to contribute TEI-encoded examples (either fragments or complete texts) that are applicable to any of these subjects. Examples are preferably encoded as TEI P5 XML texts, but also texts encoded in TEI P4 XML, other XML formats, or other (documented) electronic formats are of interest. Even examples of less-ideal encoding practices are welcome, since the idea of learning by error is a valuable didactic principle. Please do provide some indication of the errors or controversies in such examples when appropriate. After selection and editing, the example fragments will be incorporated in the freely available online deliverables, which will be issued under a Creative Commons Attribution Non-Commercial No Derivatives licence (see <http://creativecommons.org/licenses/by-nc-nd/2.5/>). All contributors will be credited.

The examples can be sent (preferably compressed in .zip format and with an indication of applicability and credits due) to teibyexample@kantl.be. Please do not hesitate to contact us for any inquiries regarding copyright issues or any more general issues.

Kind regards,

The project team:
Ron Van den Branden,
Melissa Terras,
Edward Vanhoutte

Appendix B: GANTT chart