

## AHRC ICT Methods Network Workshop

# DEVELOPMENT OF SKILLS IN ADVANCED TEXT ENCODING WITH TEI-P5

OXFORD UNIVERSITY COMPUTER SERVICES, 18 - 20 September 2006

## FINAL WORKSHOP REPORT

### 1. Introduction

It is widely recognized that text encoding -- that is, the representation of textual structures and interpretation in a portable and long lasting digital form -- constitutes an essential component in the skills portfolio of today's researcher in the arts and the humanities. Yet there is surprisingly little consensus about the best way of teaching this technique, or about how best to tailor training in it to the widely different communities needing to take advantage of it. The AHRC ICT Methods Network therefore recently funded a three-day exploratory workshop on Advanced Text Encoding Techniques at Oxford University Computing Services.

The workshop was organized by Lou Burnard, James Cummings, and Sebastian Rahtz, who play a major role in the development and maintenance of the *Guidelines for Electronic Text Encoding and Interchange* of the Text Encoding Initiative (TEI), which constitutes the major international standard in this field. The event attracted more than twenty participants from leading research and support institutions across Europe.

Its purpose was to explore the different approaches currently used in the development of text encoding skills appropriate to resource creation and analysis in the arts and humanities. The programme combined intensive practical work with refresher style presentations of the basic TEI framework, contents, and supporting technologies from the organizers. In addition, three distinguished practitioners from the US, Europe, and the UK presented contrasting approaches to the teaching of these topics: Julia Flanders (Scholarly Technology Group, Brown University) introduced syllabus options appropriate for humanists; Laurent Romary (Max Planck Institute, Berlin) demonstrated a practical way of introducing the basic architecture and technical possibilities of TEI P5 using freely available TEI web tools; and Melissa Terras (School of Library, Archive and Information Studies, University College London) gave a thought-provoking presentation on the special challenges faced by those seeking to introduce the TEI in the context of a traditional masters level humanities computing course.

The bulk of the Workshop was however devoted to exploratory practical work in small groups. Participants were invited collaboratively to explore the definition of a one-day introductory workshop on the TEI on the first day, and rapidly reached consensus on basic approaches and appropriate content. On the second and third day, each of the five groups worked on defining contents, outcomes, and appropriate materials for a range of different teaching scenarios. For each scenario the groups were asked to design the course programme and timetable, create a course synopsis, a set of learning aims and expected outcomes, and a detailed list of secondary materials including a bibliography. In a final light-hearted session, a spokesperson from each group tried to persuade a skeptical band of funders to finance their proposed approach.

All the work done was recorded using a Wiki (still available at the URL <http://wiki.oucs.ox.ac.uk/tei/>) and the materials created form the major part of this final report from the Workshop.

### 2. The generic one-day workshop

Each of our five groups worked on the same task for the first day of the workshop. Their task was to create course materials for a hypothetical one day introductory course on the TEI to be offered at some university but with registration open outside the institution. We asked each group to produce

- Full course synopsis
- Course programme and timetable
- Learning aims and outcomes

- Detailed list of secondary materials provided
- Bibliography of other related resources

The course, we suggested, should assume only basic IT literacy in the students, and should also anticipate that participants would come from a variety of higher and further education institutions, funded academic research projects, libraries, museums, and commercial companies.

As well as recommendations for topics, texts, and examples suited to such a course, we asked for course descriptions that addressed such practicalities as whether or not to involve visiting lecturers, whether students would undertake practical exercises and how these would be assessed, how the course topics would be organized and how course materials would be disseminated.

Inevitably, in the time available, groups were not able to do more than sketch out the basics of what such a course should aim to cover and how it should be organized. Interestingly, a consensus as to the key topics and outcomes for such a course emerges quite clearly from the proposals. The learning outcomes proposed by Group 1, for example, were as follows:

- Understanding value and principles of markup
- Understanding syntax of XML
- Understanding scope and philosophy of the TEI
- Ability to do simple document analysis
- Experience in creating simple TEI document
- Understand possibilities for querying and publishing TEI XML documents
- Knowledge of how to find out more about TEI
- Enthusiasm for TEI principles

The stated objectives of the course defined by Group 4 were to ensure that students would

- Understand the concept of markup (bringing intelligence to documents) and of markup as scholarly interpretation
- Be introduced to the importance of tagging, moving from visual to semantic markup
- Experience creating a TEI XML document in an XML editor
- Gain the ability to learn more about TEI, know how to get software, etc.
- Be introduced to the concept of a machine-readable grammar of rules
- Be given a sense of the kinds of things the TEI is good for.

In the topics and exercises proposed for individual parts of the day there was also a consensus about the need to combine factual information about the basic aspects of the TEI scheme with more problem-oriented and open-ended exercises focussing on document analysis. There was an evident concern that students should not simply learn the basic TEI tags parrot-style but should be encouraged to master at least the basics of customization, and to problematize the choice of markup for specific applications.

Recurrent components in each proposed course included:

- Theoretical justification and background: such topics as document structure analysis, markup theory; basics of XML; historical context and objectives of the TEI
- TEI basic structure, either based on TEI Lite or using Roma to create something simpler

- Hands-on experience using an XML editor to mark up a simple document
- Demonstration of the scalability and customisability of the TEI scheme
- Practical experience displaying a TEI/XML document.

Participants took different views on how many of these could be fitted into a one day course or to what level.

To see the five course outlines in detail, please visit the Wiki pages at <http://wiki.oucs.ox.ac.uk/tei/GroupWorkDay1>

### 3. Five workshops

We hypothesized that TEI training should be organized differently for different training contexts, and therefore asked each of our five groups to focus on one of five different learning situations or scenarios for the remainder of the workshop. The five sections constructed on the Wiki give full details of the different approaches that were judged necessary in each case: see <http://wiki.oucs.ox.ac.uk/tei/GroupWorkDay2and3>. As might be expected there was some variation in the expected Learning Outcomes for each scenario, as there was in the course contents for each scenario. Some common themes do however emerge, suggesting that it might be worth compiling a library of TEI-specific learning modules for deployment in different contexts.

Work Group 1 was asked to produce a "Teach Yourself TEI" module, which might be offered entirely online to introduce TEI to a non-specialist working alone. The expected student profile combined professional IT people working in support of TEI-related projects (e.g. project managers, document analysts, programmers, and encoders) as well as academics having to carry out similar functions on their own behalf.

Work Group 2 was asked to produce an intensive course for TEI Implementors, to be taught by a combination of lecture and practical group work. The intended audience was technically-oriented but a significant amount of course time would be devoted to interaction with non-technical specialists.

Recognizing the importance of the library community in the TEI world, Work Group 3 was asked to produce an intensive workshop aimed at practicing librarians and resource managers, who might be expected to have an interest in and some awareness of the TEI but no particular technical background.

The TEI also forms a part of a broader 'digital humanities' arena, which forms the basis of masters level courses at a number of European universities. Work Group 4 was asked to consider what aspects of the TEI should be selected for inclusion as modules within such a course.

Finally, Work Group 5 was asked to produce a workshop for a group of academic specialists (historians and archivists) who have a high degree of familiarity with their source material but comparatively low levels of IT literacy.

The following emerged as key topics, nearly all of which featured in each proposed scenario, though with varying degrees of detail:

1. Document analysis and its relevance to textual study or digital humanities
2. Markup and XML technologies
3. TEI structural elements
4. Motivation for and usage of the TEI Header
5. Hands-on experience of using XML tools, especially validating editors
6. Transformation of TEI XML using XSLT
7. Publishing of TEI XML online or in print
8. Customizing TEI, fundamentals of the ODD language and the class system
9. Coverage and scope of the TEI Guidelines overall
10. Detailed examination of some specialised TEI modules
11. Analysis of existing TEI projects: their usage of TEI and the technical solutions adopted
12. Relation of TEI to other standards or codes of practice

Participants also identified a large range of relevant teaching support materials, including many online resources, which ranged from straightforwardly didactic "Teach Yourself XML" websites to surveys of existing markup practice, and included many exemplary (or not so exemplary) sites. TEI has reached the stage in its development when it is itself the topic of research and survey, which can only improve the number, accuracy, and coverage of such resources. The TEI web site at <http://www.tei-c.org> is the natural place to look for such information, although (as one or two participants noted) this is in not currently as up to date as it might be. Projects such as [TEI By Example](#) or the [Markup Analysis Project](#) promise to offer an excellent range of learning materials, to complement the online learning materials available from several other sites, notably at Brown and Oxford.

Each of the courses developed by our participants placed a great emphasis on 'learning by doing' and on tailoring the training materials to the needs and expectations of the specific learner community being addressed. This kind of pedagogic style seems particularly appropriate to initiatives such as the TEI which have their origins in user needs, and are widely perceived as being responsive to a large user community for their development and maintenance. Whether or not, therefore, a definitive register of teaching modules for the TEI can or should be developed, it is clear that its application and usage will always rely on the skills of the individual trainer, both in engaging with the concerns of the specific learner and in tailoring the available material to suit those concerns.