Markup, TEI, Digital Libraries and Humanities Scholarship

Susan Hockey
s.hockey@ucl.ac.uk

TEI Consortium Meeting
Chicago 11 October 2002
Humanities Scholarship in the Future

- Electronic with the Web as its focus

- Use of electronic sources

- Production and “publication” of new material in electronic form

- Web becomes the digital library

Susan Hockey, TEI Meeting Chicago, October 2002
Conventional Library and the Digital Library

• Major difference
  – In the conventional library content is fixed
  – In the digital library content can be dynamic
Digital Library Initiatives: Digital Images

- Cheaper
- Easier
- Mimics non-electronic mode of working
- Fits better into conventional methods for handling material
Digital Library Initiatives: Encoded Text

• Real digital library
• Search it
• Select subsets
• “Mix and match” bits of text

• XML makes all of this possible
Humanities Source Material

• Same source used for different purposes
  – literary, linguistic, historical, cultural etc

• Same source used by different types of users
  – academic, high schools, lifelong learners, “the citizen”
Humanities Source Material: Uses of Textual Sources

• Conventional retrieval systems
  – Retrieval of complete documents – user finds all documents about a certain topic, then reads them. Often secondary sources – or perhaps historical sources

• Analysis
  – User studies content of source. Often primary sources. Literary and linguistic analyses
Humanities Source Material: Non-textual

- Photographs
- Audio
- Video and other multimedia
- Also numeric formats

- Part of the library, but less likely to be found in conventional libraries
What Does the TEI Offer for Digital Libraries?

- A framework for encoding text
- A multi-purpose encoding scheme
- A better encoding system for text than any other previous one

- But how does this really work?
- And what else do we need to do?
TEI and Libraries: Encoding

• Not a mechanical process
• Who does the encoding?
• What level of markup do they use?

• Encoding is interpretation
• Encoding drives what can be done with the document
• To what extent is this the role of a library?
Libraries and Standardization

• In the library, some standardization is essential
• But at what level?

• Is TEI lite the answer here – but does it really work?
TEI and Libraries: Delivery

- Few people will get an electronic text from a digital library and use their own software on it
- Most will want to use an existing delivery system
- But with what functionality?
TEI and Libraries: Delivery

• Retrieval – look up words and phrases
• Determined by the encoding, but also by the indexing
• Who decides what to index?
TEI and Libraries: Delivery

- Browsing
- Random movement from one piece of information to another

- Can be done by linking
- But who decides what the links are?
TEI and Libraries: Delivery

- Can automate some of the searching and linking

- But it does not always make sense
  - Same name appearing several times in one document
  - Importance of links – intellectual decision
TEI and Scholars

• Interested in their own thing
• May want to add markup with their own interpretations
• Can this be done on top of a generalized version of the document?
TEI and Scholars

- Layering of markup
- Is this different layers with different levels of complexity or importance?
- What is important to one scholar is not for another
- Is it necessary to differentiate the importance of layers?
Layered Markup – in the Digital Library

• For interpretation – linguistic, literary, historical, cultural etc
• Whose interpretation?
• How is the interpretation justified and explained?
Layered Markup – in the Digital Library

• How will another scholar use somebody else’s interpretation embedded in a document?
• Do they agree with all of it?
  – Probably not
• Do they agree with some of it?
  – Possibly
TEI and Scholars

• “Chunking” of documents
• Identify and isolate portions of documents
• Can link these together in different ways
Chunking of Documents

- Generate new “publications” automatically
- Produces something which is more database-like
- Does not read well
Chunking of Documents

- Automatic linking
- Does not always make sense
- May not want to link to every instance of something
- Some links are important for some people, other links for other people
Chunking of Documents

• “Floating annotations”
• Could be useful for some applications
  – Historical editions
  – Language learning
  – Metrical structures
• But do you link to them every time?

Susan Hockey, TEI Meeting Chicago, October 2002
Markup and Users

• Can use the markup to target different types of users
  – Dictionaries - from pocket size to large
• Perhaps also for different types of annotations
Markup and Users

• How to determine categories of users
• What is obvious to some users may need to be specified in detail for others
• Internet reaches a very wide audience
  – Much more difficult to determine audience
TEI and Other Markup Schemes / Standards

• TEI is only one of many XML-based systems used in humanities information
  – EAD – very widely used in archives
  – METS – metadata structures
  – Various DTDs, schemas used in publishing
TEI and Other Markup Schemes / Standards

- What is the relationship of TEI to these?
- Can all of these co-exist
- Overlap in some areas, particularly metadata structures
TEI and Other Markup Schemes / Standards

• Making different schemes co-exist
• Technical structures need to be resolved
• But content is more important for the user
• Need agreement over the meaning of the content

Susan Hockey, TEI Meeting Chicago, October 2002
Future of the TEI: Usability

- Reach out to new communities
- Need to make it easy for them to understand what this is about
- Easy to do the markup
Future of the TEI: Tools

- Windows-based
- Easy to use
- Tools to do the markup
- Tools for delivery and analysis
  - Need much more research on functionality
Future of the TEI: Tools

- Concentrate on new applications, not mimicking print or only retrievals
- Moving to a world of multimedia
- Convergence of TV and Internet
- Encoded text will be the wrapper for all of this

Susan Hockey, TEI Meeting Chicago, October 2002
TEI and Digital Scholarship in the Future

• Research
  – on the uses and users of digital resources
    • who are the users?
    • what functionality do they need?
  – towards reducing the cost of markup
    • can markup be done cheaper?
    • where is the trade-off between, quality, cost and getting a project finished?
TEI and Digital Scholarship in the Future

• Co-existence with other standards
  – Work on harmonization

• Education of information professionals, users, publishers in order to
  – see each other’s perspectives
  – recognize each other’s viewpoint