



Markup Issues

Stephen Miller

AAC / ÖAW

Texts & Files SS 2007

Universität Wien



They shall be nameless...

- *Encoding Guidelines*

“Generally speaking, if something is there on the page as part of the text, include it in the electronic transcription, using the appropriate tag or tags to encode it.”



How much tagging is
enough?



“Thomas Kellys and Famalyys Journal” (1827)

Started From Doolough in the parish
of Jurby in the Isle of Mann on the 3rd
of July for to Take our passage for
Liverpool Engaged with the master of
th John Bull of Ramsey for our
passage at 5s a passenger our
Famaly consisted of Thomas Kelly
Senior Thos Kelly Junior



What do I need to tag? Placenames?

Started From **Doolough** in the parish of **Jurby** in the **Isle of Mann** on the 3rd of July for to Take our passage for **Liverpool** Engaged with the master of th John Bull of **Ramsey** for our passage at 5s a passenger our Famaly consisted of Thomas Kelly Senior Thos Kelly Junior



What do I need to tag? Names of vessels?

Started From Doolough in the parish of Jurby in the Isle of Mann on the 3rd of July for to Take our passage for Liverpool Engaged with the master of th **John Bull** of Ramsey for our passage at 5s a passenger our Famaly consisted of Thomas Kelly Senior Thos Kelly Junior



What do I need to tag? Editorial interventions?

Started From Doolough in the parish
of Jurby in the Isle of Mann on the 3rd
of July for to Take our passage for
Liverpool Engaged with the master of
th[e] John Bull of Ramsey for our
passage at 5s a passenger our
Famaly consisted of Thomas Kelly
Senior Thos Kelly Junior



What do I need to tag? Individuals?

Started From Doolough in the parish of Jurby in the Isle of Mann on the 3rd of July for to Take our passage for Liverpool Engaged with the master of th[e] John Bull of Ramsey for our passage at 5s a passenger our Famaly consisted of **Thomas Kelly Senior Thos Kelly Junior**



What do I need to tag?

and Wife Isabel Kelly Daughter of Thos Kelly
Senr Thos Kelly Juniors Children Jane Ann
Margaret Isabella and Maria and also John
Quayle Ann Quay Christiana Tear Beside
John Cannells Famaly John Canell and Wife
his Children John Ann William and Thomas
and Isabella Callister Niece to Thos Kelly
Senier

**This is only the *first* entry in the
Journal**



TEI Guidelines

- The TEI Guidelines are *descriptive* and not *prescriptive*
- What is to be encoded (and equally as importantly) what is *not* to be encoded depends on the aims of the project



Encoding: Best Practice

Stephen Miller

AAC / ÖAW

Texts & Files SS 2007



Encoding: Best Practice(s)

- Digital Library Federation
 - *TEI Text Encoding in Libraries: Guidelines for Best Encoding Practices 2.1* (16 November 2005)
- University of Wisconsin-Madison
 - *Guidelines for Markup of Electronic Texts 2.0* (27 March 2006)



University of Wisconsin-Madison

- University of Wisconsin-Madison
 - *Guidelines for Markup of Electronic Texts 2.0 (2006)*
 - <http://uwdcc.library.wisc.edu/resources/etext/TEIGuidelines.shtml>



5 Levels of Encoding

1. E-Facsimile Level
2. Reading Level
3. Pedagogical Level
4. Scholarly Level
5. Dictionaries



Level 1

- E-Facsimile Level

- This level is intended to support e-facsimile collections with no electronic text beyond OCR and metadata. The markup used for these collections is automatically generated, and is based on the data model developed for those collections [UWEFacs]. That specification is not within the scope of this document



Level 2

- Reading Level

- Texts at this level need to be marked up at a level minimally sufficient to support basic reading, browsing, retrieval, and navigation. The elements and attributes used at this level are, therefore, only those needed to format a text coherently on a screen.



Level 3

- Pedagogical Level
 - Markup at this level should encode structure sufficient to enable search, retrieval, and display for the purposes of teaching or basic research in support of a subject or discipline. It may also provide for additional retrieval and display options as needed for purposes of text analysis. Additionally, it may contain references to external documents for purposes of text comparison, alignment, and reference.



Level 4

- Scholarly Level

- The markup used for this level will be determined according to the needs of the researcher. Hence, no specific guidelines will be developed for Scholarly Level collections. Still, to the extent possible scholars should make every effort to follow the various guidelines defined for the other levels. Doing so will simplify the task of indexing and interface development and should also provide users with a more consistent environment.



Digital Library Federation

- Digital Library Federation

- *TEI Text Encoding in Libraries: Guidelines for Best Encoding Practices 2.1* (2006)
- <http://www.diglib.org/standards/tei.htm>



5 Levels

1. Fully Automated Conversion and Encoding
2. Minimal Encoding Purpose
3. Simple Analysis Purpose
4. Basic Content Analysis
5. Scholarly Encoding Projects



Level 1

- Fully Automated Conversion and Encoding
 - Purpose: To create electronic text with the primary purpose of keyword searching and linking to page images. The primary advantage in using the TEILite DTD at this level is that a TEI Header is attached to the text file.



Level 2

- Minimal Encoding Purpose
 - To create electronic text for keyword searching, linking to page images, and identifying simple structural hierarchy to improve navigation.



Level 3

- Simple Analysis Purpose
 - To create text that can stand alone as electronic text and identifies hierarchy and typography without content analysis being of primary importance.



Level 4

- Basic Content Analysis

- To create text that can stand alone as electronic text, identifies hierarchy and typography, specifies function of textual and structural elements, and describes the nature of the content and not merely its appearance. This level is not meant to encode or identify all structural, semantic or bibliographic features of the text.



Level 5

- Scholarly Encoding Projects
 - Level 5 texts are those that require subject knowledge, and encode semantic, linguistic, prosodic or other elements beyond a basic structural level.