

Open E-book vs. TEI-lite; ebook standards compared

John A. Lehman
Professor of Information Systems and
International Business
School of Management
University of Alaska Fairbanks
John.lehman@uaf.edu

Abstract

This paper compares two leading XML markup sets for electronic book publishing. TEI-lite has for several years been the default markup for scholarly text archive projects. Open-Ebook is an industry-developed standard currently championed by Microsoft. This paper examines the two in terms of their views of document architecture, metadata, and implementation. It also examines issues in the automatic conversion of documents marked up with TEI-lite to Open-Ebook.

academic community

- standardize on SGML/XML-based markup
- development of a common document type definition (DTD) for scholarly texts
 - Text Encoding Initiative (TEI)
 - TEI-lite

Publishing industry

- standardize on XML-based markup
- development of a common DTD: Open Ebook (OEB)

Questions:

- Whether TEI-lite could be replaced by Open Ebook as the standard DTD for scholarly text archives?
- Whether TEI-lite documents can be mechanically translated to Open Ebook documents?
- If TEI-lite documents can be mechanically translated to Open Ebook documents, what is lost in the translation?

Architecture of TEI-lite

- designed to preserve structure of underlying texts
- designed to preserve knowledge about underlying texts
- The objects represented are generalized from manuscripts and printed books
- preservation of differences between different instances of an underlying text critical
- not designed as a means of presentation

Architecture of Open-Ebook

- designed to standardized the publication of modern electronic books
- three components:
- a package for the components of an electronic book,
 - facilities for marking up structure,
 - markup to aid automatic screen-based typesetting
- maintains compatibility with HTML where possible.

Metadata in TEI-lite

packaged in the same file as the text, in the form of a header

\. A detailed example of the metadata available is given in Appendix A.

Metadata in Open-Ebook

- contained in separate files within the Open Ebook package
- The parts of the package file are:
 - **PACKAGE IDENTITY** – a unique identifier for the OEB publication as a whole.
 - **METADATA** – Publication metadata (title, author, publisher, etc.). Dublin Core
 - **MANIFEST** – A list of files (documents, images, style sheets, etc.) that make up the publication. The manifest also includes fallback declarations for files of types not supported by this specification.
 - **SPINE** – An arrangement of documents providing a linear reading order.
 - **TOURS** – A set of alternate reading sequences through the publication, such as selective views for various reading purposes, reader expertise levels, etc.
 - **GUIDE** – A set of references to fundamental structural features of the publication, such as table of contents, foreword, bibliography, etc.

Discussion

- original goals for the TEI-lite and the Open Ebook DTD's were very different
- the architecture of DTD is derived from its goals
- each is incapable of representing a substantial subset of the contents of the other

Can Open Ebook can serve as a publishing format for texts marked up in TEI-lite?

- *Reading* texts in TEI requires the construction of a stylesheet (generally using XSLT) to translate the TEI-lite markup into HTML.
- Open Ebook is designed to represent printed books electronically
- ,provides a very convenient format for electronic text publication
- Tools currently used to translate to HTML may be used with minimal changes to translate to Open Ebook