

The Text Encoding Initiative and Electronic Legal Texts

by Joe Thomas

1-3

In "The Text Encoding Initiative and Electronic Legal Texts" program, **Kevin Butterfield** (Technical Services Librarian at Southern Illinois University Law Library) and **Nicholas D. Finke** (Head of Library Publications and Director of the Center for Electronic Text in the Law at the University of Cincinnati College of Law) guided participants through a thorough introduction to the Text Encoding Initiative (TEI). Finke led off the presentation with an overview of *text markup*—that is, encoding text within a document based on the type and purpose of the text. TEI goes beyond HTML (used primarily to *present* text in a particular way) by indicating the structure and meaning of text. TEI was founded in 1987 to develop a standard for academic text markup. The goal of the initiative is to set a standard for text encoding in order to create "Document Type Definitions" (DTDs) that conform generally to the standard but which are still specific to particular areas

(e.g., humanities, social sciences, or law). Use of a DTD system for text mark up offers a set of rules for describing a document, including its structure, and expressing that

in a standardized way. This is important most obviously to a library when it is acting as an electronic publisher, but it is also important to know the standard for ongoing access to documents in these formats. There is also a preservation issue involved. Documents written under TEI guidelines are encoded for future use in a logical and standardized way. TEI specifically offers the advantages of training, good software support, the ability to handle legal text, sophisticated cross-reference models, and easy extensibility.

Butterfield then presented an analysis of the basic structure of a TEI document. He first described the basic functions and capabilities of TEI coding in brief: it describes bibliographic and non-bibliographic information; it supports the identification and retrieval of the text; and it supports machine analysis of the encoded text. Fortunately for librarians, the formulation of TEI standards relied heavily on existing standards, such as MARC and AACR2. A TEI document is first broken into two parts: the text of the document itself, which can then be broken down into many additional levels of complexity, and the header. The header is the necessary element in a TEI document and remains an integral part of the document. The header contains the metadata that identifies the document with great specificity. The four elements of the header are 1) file description; 2) encoding description; 3) profile description; and 4) revision description. Elements two through four are relatively straightforward. The first element, file description, can be broken into other elements, three of which are mandatory: title statement, publication statement, and source description. This element acts as the document's "title page." Although TEI data can be created to be AACR2-conformant, there are no requirements to that effect. This makes mapping TEI document headers to MARC technically possible, but also a process that will require significant local checking.

Both speakers made it clear that increasing familiarity with and utilization of TEI standards will be necessary for the work of law librarians in the not-too-distant future.

Joe Thomas (*thomas.2@nd.edu*) is Head of Technical Services at Notre Dame Law School, in Notre Dame, Indiana.

Order the audiotape of this (or any other) program! Call Mobiltape at 800/369-5718 or order online at <http://www.mobiltape.com>.