Data Metastructures: From Style Guides to Structure Guides

Moderator:
Peter Goldie
Lightbinders Inc
San Francisco, California

Panelists:
Robert H Dolin
Kaiser Permanente
Oakland, California

Stuart Barton
Clinical Evidence
London, UK

Reporter:
Teresa Hood
Centers for Disease Control
and Prevention
Atlanta, Georgia

In a brief introduction, Peter Goldie outlined several areas of text structuring with metadata whose importance became evident during the Council's 1999 conference: the necessity of tagging early in production, elements and information architecture to consider before tagging, and the increased use of extensible markup language (XML).

Robert H Dolin sits on the Health Level 7 (HL7) committee in an American National Standards Institute–accredited organization working to develop standards for health-care communication protocols. HL7 specifies standards for reporting of patient records with XML, which are aimed at improving messaging between health-care systems and aiding researchers seeking data. It focuses on clinical information, but many of its standards could be used to create guidelines for structured reporting of research or journal articles. Other standards and drafts that could be used for journal text are

1. PubMed record field descriptions (*www.ncbi.nlm.nih.gov:80/entrez/query /static/help/pmhelp.html #SearchFieldDescriptionsandTags*).
2. PubMed SGML DTD (*www.ncbi .nlm.nih.gov:80/entrez/query/static/spec .html*).
3. InterMed's GuideLine Interchange Format (GLIF) (*www.glif.org*).
4. Clinical Data Interchange Standards Clinical Trials DTD (*www.cdisc.org*).
5. Dublin Core MetaData Project: Metadata for Electronic Resources (*www.purl.org/metadata/dublin_core/*).

PubMed's document type definition (DTD) and defined tags are probably the closest thing to an existing standard for tagging and reporting journal text. Existing terminologies that could be used as standard syntax in structuring journal articles include MeSH, SNOMED, and LOINC.

Dolin stressed the importance of identifying all possible information needs before building a structural model so that code and information can work fluidly in tandem. Information requirements should come before exact XML styles are decided. HL7's information model for patient record architecture (PRA) was outlined as an example of a structural model. Finally, Dolin presented a journal article tagged by using a combination of XL7's PRA DTD and PubMed's DTD, clearly illustrating how metadata applied to table and text information aid in its retrievability.

Stuart Barton spoke next of actual experience in tagging information for the British journal *Clinical Evidence (CE)* (*www.clinicalevidence.org*), where he is a clinical editor and serves as head of research and development. Published every 6 months, *CE* differs from other medical journals in that editors first identify a question likely to be posed by doctors or patients, then summarize the available information that best answers the question. Of the questions posed, 80% are successfully answered; unanswerable questions are identified as such.

CE's management feels that structuring text as data is an investment that gives returns by adding efficiency, diversity, and adaptability to information retrieval. Such ease of retrieval might be necessary for future survival; research has shown that patients and physicians routinely go to sources from which they have a high expectation of quickly finding relevant information—those with a high value as shown by the equation

Usefulness = RelevanceX(Validity/ Effort).

Barton said that experience with *CE* has shown standard information structure to be less important than actual tagging with some sort of systematic approach. That observation was based on the vast amount of necessary tagging (even though *CE* tags only tabular data) and on research showing that information search and retrieval is less linear than imagined. Staff should identify what information will be tagged, how tagging should be done, and who will do the tagging. After identifying the minimal amount of structure necessary for expected needs, journal staff should just start applying metadata to their information. For unidentified future needs, more tagging rather than less might be a safer option. Barton summed up *CE*'s approach to metatagging of information as, "Agree on the aims and allow the structures to evolve."

During the discussion after the presentations, a question was posed concerning the relationship between data structures and accepted styles, especially in view of the Council's revision of its style manual. The speakers agreed that it would be shortsighted of the Council not to consider data structures in revising the manual, given the current and future importance of metadata in publishing.