

The ABCs of Data Conversion

Chair:

Eli Willner

Data Conversion Laboratory
Fresh Meadows, New York

Presenters:

Richard A Wiklund Jr

Cadmus Journal Services
Baltimore, Maryland

Robert Kendall Oberteuffer

Johns Hopkins Medicine Center for Information
Services
Baltimore, Maryland

Reporter:

Lori Cagle

Texas A&M University
College Station, Texas

If successful relationships require good communication, perhaps a friendlier bond can be forged with computers if people learn to speak their language. Receiving information in a language unknown to a particular computer is a common source of frustration in this electronic age. Data conversion, the ability to translate information into other forms, has become a valuable skill.

Richard A Wiklund Jr explained data conversion as the mapping out of steps to convert an original document (origin) to a desired end product (target). Some examples of the process are converting WordPerfect into Microsoft Word or SGML into HTML. The “save as” function is a data-conversion tool. Custom tools can also be developed in the form of macros.

Wiklund described some of the most common text and graphics targets:

- PDF—Portable Document Format. A conversion of the electronic typeset page retaining all fonts and layouts.
- ASCII—pronounced ask´-ee. The lowest common denominator in data formats

commonly used on the Internet because of its accessibility to many different computers. When saving a document as text, choose ASCII as the target format. No fonts, colors, or point sizes are converted.

- SGML—Standard Generalized Markup Language. A combination of a coding structure and ASCII. Coding structures are called document type definitions (DTDs) and are used to customize ASCII. An analogy: different colors may be added to a paint base to get custom shades; the DTD tells what “colors” should be added to ASCII to reach the target of SGML. Can require several steps.
- HTML—Hypertext Markup Language. A basic DTD for SGML. Often requires manual intervention to reach desired end.
- XML—eXtensible Markup Language. New technology that streamlines SGML so that it can be used for online publication.

Wiklund said that the goal of data conversion is data normalization—changing original material in various forms to one standard format. Although the process adds expense, it also provides opportunities. Conversion programs customized to a journal style give editors more time for substantive editing. Normalized data can be marketed to other companies, such as OVID, or used to create indexes of authors, subjects, or nearly anything else.

Following Wiklund’s answers to what and why, Eli Willner explained how. Strategies for data conversion, Willner said, usually must involve both software and human intervention. The key to successful conversion is to plan the steps carefully.

Willner presented the following steps:

- Step 1—Know your data. The amount of data, the original and target formats, text and cross-reference complexity, and type and quality of images must all be analyzed to determine costs, schedules,

software needs, and portions that require human tweaking.

- Step 2—Decide whether to do it in-house or outsource it. Advantages of doing the project internally include total control, absence of security issues, schedule flexibility, and ease of making changes. Advantages of contracting the project out are levels of expertise, low drain on internal resources, lower cost, increased accuracy and timeliness, and early identification of major problems.

- Step 3—Choose a DTD. Two options are industry standards, such as OVID, and government-funded public-domain DTDs. Custom DTDs are expensive to develop, but semicustom DTDs offer a happy medium by custom-tailoring standard DTDs to suit a particular project.

- Step 4—Choose the right combination of human and machine to complete the project with maximal quality and minimal expense.

- Steps 5 and 6—Develop a conversion schedule and create a conversion specification that maps out in writing exactly how the converted data will look.

Other planning considerations that Willner noted are quality assurance, implementation, and staff education.

Sharing his experience with data conversion, Robert Kendall Oberteuffer presented a case study of an SGML-to-HTML conversion. The project was the electronic publication of *Annals of Thoracic Surgery*, a monthly journal of 100 to 120 articles, each 2 to 7 pages long, with moderate use of graphics. The project was handled internally, and Oberteuffer discussed project goals, conversion options and approaches, tools (software) and codes used, stages of development, handling of special characters, and publication steps. He also listed some Internet resources for people interested in trying their own data-conversion project. 