
Mikhail Vaysbukh

**Leveling The
Textbook Playing
Field For The Print
Disabled**

Data Conversion Laboratory Inc.
www.dclab.com

ISBN: 1-4116-0056-8

Headquarters

Data Conversion Laboratory Inc.
184-13 Horace Harding Expwy
Fresh Meadows, NY 11365
Tel: 718-357-8700
Fax: 718-357-8776

World Wide Web

<http://www.dclab.com>

Copyright© 2003 Data Conversion Laboratory Inc.

All rights reserved

Contents

LEVELING THE TEXTBOOK PLAYING FIELD FOR THE PRINT DISABLED... 2

ACCESSIBILITY ACT SLOW TO GO THROUGH SENATE	2
WHAT DOES IT TAKE TO CONVERT TEXTBOOKS TO XML?	3
UNDERSTAND THE ISSUES FIRST	4
ALL ELEMENTS NEED TO BE DEFINED	4
LEVEL OF GRANULARITY	6
SO WHAT ARE YOUR OPTIONS?	6
TRANSFORMING THE LIVES OF THE PRINT DISABLED	8

Leveling The Textbook Playing Field For The Print Disabled

By Mikhail Vaysbukh

[Data Conversion Laboratory](#)

Increasing numbers of state legislatures require textbooks and other educational materials to be in structured electronic formats like XML, which makes educational texts more accessible and vastly improves the quality of life for people who are not able to read due to disabilities. This white paper looks at current accessibility legislation and reviews many of the issues surrounding the conversion of materials to meet these legal requirements. It also provides an overview of how a conversion firm, Data Conversion Laboratory, goes about converting data into XML and other structured electronic formats.

More and more states in America are requesting that textbooks and other learning materials are available in structured electronic format to improve accessibility for people with disabilities. In early 2004, for example, Kentucky's state legislature will require all educational materials be in well-formed HTML or in XML files that meet accessibility requirements.

There are many advantages to having educational materials in electronic form. It simplifies translation to Braille. Lets students with low vision see text on a monitor screen, not just hear it. And allows advanced navigation through documents by automatically extracting document hierarchies.

When electronic material is structured in well-formed HTML or in XML, it offers direct benefits to people with print disabilities. When structured data is read by a text-to-speech engine, for example, it can supply crucial information like what page or heading level you are on. It also helps generate a navigational structure, which enables the user to move from level to level and be able to recognize things like the start of the next exercise.

Accessibility Act slow to go through Senate

Besides state legislation in the United States, there is also the Instructional Materials Accessibility Act (IMMA), a Federal initiative currently going through the Senate as a new bill. The accessibility rules written into the Act are similar to

those of state legislatures.

Unfortunately, the IMMA is moving somewhat slowly through Senate, which is why many individual states decided not to wait and came up with their own legislation. That legislation is, of course, only valid on a State level. Whereas the IMMA would be at Federal level and would cover all states.

If implemented, the IMMA would call for a national XML standard for accessibility. This would require the creation of a technical panel to select or create such a standard.

Interestingly, the Department of Education has awarded CAST (Center for Applied Special Technology, www.cast.org) a grant to set up such a panel -- even though no accessibility law has yet been passed in Senate. Such a panel, however, would only have the power to recommend an XML standard and suggest that publishers comply with it on a voluntary basis.

Despite the delays-to-date in the Senate over national accessibility legislation, it is clear that it will go through at some point. And already an increasing number of states require by law that educational materials be accessible, which normally means they must be in well-formed HTML or XML.

For many publishers XML is preferred over HTML because it separates content from format and allows you to publish in print, on the web, CD-ROM, or PDA, at the click of a button.

What does it take to convert textbooks to XML?

XML specifications that have been put forward for a National Standard for Accessibility include: Clearly defined document hierarchies, presence of a text description of images, and table structures that facilitate import into tools for the visually impaired (like defining table headings). These will help people who are blind. But the two most important concerns for the publisher become quality and cost. Clearly, reputable publishers do not want the quality of their material to suffer during the conversion process, and want accessible titles to be well received both by the accessibility community and by the media. Therefore a good deal of attention needs to be paid to the quality of the conversion.

At the same time, publishers want to take the least labor intensive and most cost-effective route, which means automating much of the process. The other consideration is scalability. The conversion process put in place needs to work with future publications too - otherwise it would be untenable.

Understand the issues first

Before going ahead with a conversion project, you need to get a full understanding of your situation and the issues surrounding it. Data conversion is not as simple as it sounds. There are no magic buttons that will convert Quark to XML or Miles to XML, for example. Your publishing files will likely have missing data that will need to be inferred so it can be accurately converted. This usually has to be done by editors (but not always -- the conversion software designed by Data Conversion Laboratory, for example, automatically infers what a lot of software requires people do by hand). Ambiguities will need to be resolved. And there will be data that doesn't fit the Document Type Definition.

(A Document Type Definition, or DTD, is a template that enforces the structure of documents).

You may also need to impose structure where none exists, such as defining styles like "Heading 1" to all headings and "Heading 2" to all sub-headings.

All elements need to be defined

In fact, it is necessary to define every element in your files, so each can be recognized and properly converted by the conversion software. Elements that need to be defined, or "tagged," include: paragraphs, cross-references/linking, lists, tables, graphics, index, footnotes, special characters, Math, table of contents, and front and back matter.

This might seem frustrating and you might just want to "get on with it." But there is good reason to map everything out in precise detail. Electronic books - whether HTML-based, PDF, or digital talking books - are very different from printed books.

References

References are a good example. In printed books it is fine to use variations like:

- See Figure 15.5
- See fig. 15.5
- Refer to figure below
- As illustrated on previous page

But for these to work in electronic format, and for the conversion process to be properly automated, all references need to be consistent and point to exact locations in your document.

Tables

Tables are another element that needs to be consistent for the conversion process to run smoothly. The problem with tables is they aren't always created with a table editor, such as is found in Microsoft Word and other leading word processors. Simple tabs (tablature), for instance, can be used to make data look like it is in a table. But technically this would not a table. Quark Xpress doesn't have a table editor. A standalone table editor is available for Quark. But this only improves the situation marginally when converting files to XML. The reason for this is there is still no "off-the-shelf" software that perfectly converts tables into XML. Plus some third party table editors don't allow the exporting software to access the table coding.

When tables aren't built with a dedicated table editor, there is little choice but to guess what constitutes a column, row, or table spanning. And in publishing files, large tables that span more than one page are often split into more than one table, although logically they are still one table. This makes it more difficult for conversion software to identify where the table starts and ends.

Lists vs. Question Sets

Lists can also be ambiguous. For example, in textbooks it is necessary to define whether the data is actually a list or whether it is a question set. The problem is, conversion software will not be able to tell the difference unless it is given guidance (you would need a human to read it to be able to distinguish between a list and a question set). A question with multiple answers (like the example below) will look like a nested list to a computer. So it has to be differentiated in some way from a real list.

What is the capital of Great Britain?

- New York
- Paris
- London

Graphics

Graphics do not pose major problems. But to meet accessibility requirements you do need to provide a text description of the image using a similar tag to the "alternative text" attribute found in HTML.

Level of granularity

The next step is to work out how much "granularity," or detail, you want available in the converted document. For example, the references at the end of a book can be tagged as a regular paragraph. But this would mean that all the information making up a reference would be available as one chunk, not as individual components such as the author's first and last names and the publication name and date. This is fine if you just want to stylize the reference in italics. But you might prefer to markup the individual details of a reference - author's first and last names, and so on - so they can be pulled out and deployed for specific purposes.

All the above issues need to be resolved before the conversion process begins (a specialist firm would be well placed to help out with this).

So what are your options?

There are a number of options for getting your publishing files converted to XML. You can either do it yourself, outsource it to a compositor or web-development firm, or to a specialist data conversion house.

Let's look at these options in more detail:

In-house conversion

Doing the conversion in-house involves using existing technical staff or hiring new staff. You would also need to assess the programming tools on the market that manage conversion to XML (some are free). These include: XSLT (www.oasis-open.org), which is good for converting from markup languages; and Omnimark, which is powerful but complex to use. If you currently use Quark Xpress, another alternative is to use one of the third party add-on products that integrate with Quark to convert files to XML. On the surface, this may sound like the ideal solution. Unfortunately, the quality of the XML produced is not robust

enough for most people's needs and requires significant tweaking and adjustment in order to be usable. There are also a number of standalone tools that convert publishing files into XML. These include Filtrix (Blueberry), TagWrite (Zandor), and Worx (Hypervision). Details of these and similar tools can be found at www.xmlsoftware.com and <http://xml.coverpages.org>.

Outsourced conversion

Compositors:

Besides creating publishing files in Quark Xpress, Miles, or other formats, many compositors now offer some level of conversion services. If you already use a compositor, it may make sense to go down this road.

Web/multi-media firms:

These also provide conversion services as part of their package. If you already use such a firm to host electronic books on the Internet or to create books on CD-ROM, you may wish to add conversion services to the list.

NOTE: A possible downside of contracting your conversion needs to a compositor or web/multimedia firm is that conversion is not their primary business and they may not be familiar with all the potential issues. This is something you would need to evaluate individually. A further downside of using a compositor is this: If you use more than one firm (for different titles), you could well end up with inconsistent results. This is because compositors all use their own variation of XML.

Data conversion houses:

Since data conversion firms specialize in conversion, they have more experience with the various issues that may come up. It is their core business. So you can expect a very high level of quality -- particularly important for textbooks and educational materials, which typically have a lot of complexity in terms of tables, question sets, lists, and special characters.

Transforming the lives of the print disabled

While meeting the requirements of accessibility legislation is clearly essential - otherwise you would be contravening state (and soon Federal) laws – the human angle might be more important. By converting your publishing files to XML you are at the forefront of a technology revolution that is transforming the lives of people with print disabilities. As a publisher you are at the heart of this revolution.

Mikhail Vaysbukh
Director, Educational Publishing Division
Data Conversion Laboratory Inc.
Phone: (718) 357-8700 Ext.#228
E-mail: mvaysbukh@dclab.com

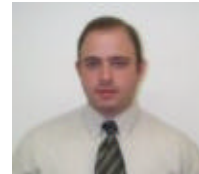
TECHNOLOGY JOURNAL

Subscribe to DCLnews for the latest on XML, data conversion, accessibility, and e-books.

www.dclab.com/dclnews.asp

ABOUT MIKHAIL VAYSBUKH

Mikhail Vaysbukh is Director of Educational Publishing for Data Conversion Laboratory Inc., Fresh Meadows, N.Y. He has overall responsibility for DCL's Educational Publishing division, which includes project management, management of the project engineering process and client development.



Mikhail is an active member in several technical industry groups such as the Open eBook Forum Publication Structure Working Group and the AAP Enabling Technology Committee. Mikhail is a co-author of the Open eBook Publication structure specification.

Mikhail has been with DCL for over 6 years and has served in a number of positions ranging from Production Manager and Project Manager to his current position of Director of Educational Publishing. He holds a BS in Computer Science from Touro College.