Adobe® FrameMaker®
Easing the transition from unstructured content to structured XML

About TEEX
As one of the largest providers of workforce training in the nation since 1919, the Texas Engineering Extension Service (TEEX) is known for its innovative, customized programs, including hands-on and on-site training. TEEX is an agency of the State of the Texas and part of the Texas A&M University System. It is headquartered in College Station, Texas with several regional training centers and offices located around the state. These facilities support six agency business units, all of which offer workforce training in particular subject areas including law enforcement, emergency medical and fire services, homeland security, and public infrastructure.

Although TEEX is an agency within the Texas A&M University System, 90% of its training is delivered outside the State of Texas. The agency is recognized as a national leader in broad-based technical training and assistance in a diverse range of response disciplines. Through more than 800 training courses, TEEX training solutions help approximately 200,000 workers each year to upgrade their skills, become certified in their professions, or learn more about new technologies.

To support these numerous training programs in a wide variety of disciplines, TEEX produces hundreds of different training manuals and workbooks each year. All curriculum is developed within the agency, and each business unit supports its own curriculum development teams who work with subject matter experts to create the training materials.

Starting point: successful process, decentralized control of formatting
Over the past six years, TEEX invested millions of dollars in a comprehensive overhaul of curriculum creation standards and processes. The agency implemented a standardized three-phase development process that provides excellent communication and feedback throughout the development lifecycle. It also provides a mechanism to ensure that the content conforms to the quality standards that set TEEX apart as a training provider. This effort has produced thousands of high-quality training manuals.

Because creating templates in Adobe FrameMaker is relatively easy, each of the TEEX business units was allowed to retain its own template. This resulted in significant duplication of effort as well as more advanced FrameMaker expertise being required by each curriculum group. This also led to divergent formatting based on personal preferences of the individuals maintaining the template in each group. In some cases, multiple copies of content were maintained to produce an instructor guide and participant manual for each course—a significant challenge with the hundreds of TEEX courses delivered annually.

It became clear that a solution that simply allowed each curriculum group to dictate formatting would not be manageable given the current level of resources. Formatting decisions had to be made agency-wide. This would free the curriculum teams to focus on content rather than concerning themselves with formatting, and the agency would also benefit by having unified formatting across all business units. Structured FrameMaker automatically enforces correct formatting, rather than relying on people to review and find formatting inconsistencies.
Choice: custom or standards-based solution

After researching multiple standardized structures, TEEX decided to create a custom solution rather than implement a standards-based solution like DITA (Darwin Information Typing Architecture). The major factors that led the agency to this decision were:

- The narrative nature of its content did not lend itself well to any current standards.
- Easing the transition for content creators as much as possible. With a custom solution, much of the complexity involved is hidden from the end user, thereby decreasing training costs.
- No current need to exchange content with other businesses.

While current business requirements dictated a custom solution, TEEX was careful to implement a solution that would allow for a simple transformation in the future if requirements change. The agency kept the structure DITA-like because DITA is the most likely standard for future implementation.

Move to centralized template and XML workflow

Moving from a situation where each business unit enjoyed a large amount of autonomy to a centrally-managed solution was a major shift for TEEX. One of the big concerns was that the business units would resist this move to a more centrally-managed solution. Involving the business units with the decision-making process through a representative group provided an excellent way to explain the benefits of relinquishing some amount of control over content formatting.

Working group

To bring consensus, TEEX formed a working group composed of instructional designers from each of the agency business units. This group did a significant amount of content analysis to determine common items that existed across the curriculum development groups—a lengthy undertaking due to the wide array of content topics and intended audiences.

Once common items were identified formatting was discussed and a single recommendation was agreed upon by the group. The ability to come to consensus decisions on almost every aspect of the template was a major accomplishment and presented a major shift in the way the agency would approach curriculum development: making formatting decisions and then removing many aspects of formatting from the day-to-day process of content creation.

The formatting recommendations were hard-coded in the Element Definition Document (EDD). By hard-coding information in the EDD rather than calling out paragraph tags, TEEX is able to remove the possibility of technical writers inserting overrides based on their own preferences. Once content exists in XML, any overrides created by the technical writer are removed when the document is saved and control of formatting is centralized, decreasing management cost.

Previously, each business unit had one individual responsible for maintaining its FrameMaker template. "Because we have centralized the formatting information, I am able to single-handedly manage the technical aspects of the template," says Dustin Vaughn, the system analyst in charge of the technical implementation. "Moving to a centralized template will significantly lessen the burden on each business unit to make formatting decisions."

The working group also provided an excellent channel of communication with the remainder of the curriculum development teams. Throughout the process, each group was made aware of progress and knew that they had a representative to voice concerns and make sure that their perspective was taken into consideration.

Another benefit of the working group was the fact that the working group members became the FrameMaker experts within their particular group. This deep understanding allows each group to have an in-house resource when they have questions and also lessens support costs for the agency by each group handling initial tech support and only escalating issues that may require additional support.
Knowledge transfer is key
Vaughn relied heavily on the expertise of Tom Aldous for training and guidance as he implemented the solution. “Our association with Tom has been one of the most significant factors in getting this fully-customized solution from white board to reality,” he explains. “Tom is not a guy who gets surprised. He knows FrameMaker and XML inside and out, and he was able to turn every crazy idea that we had into a workable solution.”

Style-based template
Based on feedback from the curriculum groups in the agency business units, the possibility of continued use of Microsoft Word as a development tool was identified as a requirement of any proposed solution. However, integrating this into a structured XML workflow presented many technical challenges because Word and unstructured FrameMaker handle many of the common content creation tasks in different ways. In addition, current workflow was completely open and allowed each content creator to format the document to fit their own requirements.

To improve the current process, a styles-based template was introduced to bring consistency and uniformity to the early content creation phase. Removing the need for the subject matter expert and the instructional designer to worry about formatting enables them to focus on the task at hand—creating content. Uniform formatting also saves the technical writer’s time when bringing the content into structured FrameMaker because there is no longer the need to guess the intent of the instructional designer and subject matter expert. The template was designed to be simple to use. "The goal of the template is ease of use," Vaughn adds. "I wanted to be able to get a new user up and running—creating content, with a minimal amount of training.”

Content conversion
Content is created by the instructional designer and subject matter expert. This process may also include the technical writer, the curriculum manager, as well as other individuals within the organization. Once the module’s text based content has been tagged with the appropriate styles, the content must be brought into FrameMaker in a structured format.

The initial step is to save the document in Maker Interchange Format (MIF). Documents are saved in the MIF format to strip out any hidden objects. The MIF document is first structured with a simple conversion table that maps style names to elements. Some simple nesting is also done to allow context to be identified, but mostly just a one-to-one mapping. Nesting at this stage is done only to make the context of the particular element available for later conversion.

In FrameMaker’s conversion table, using RE: ensures that the entire document is wrapped in a root element.

<table>
<thead>
<tr>
<th>RE:</th>
<th>Root</th>
</tr>
</thead>
<tbody>
<tr>
<td>NoName</td>
<td>Patient Packaging</td>
</tr>
<tr>
<td>ModuleTitle</td>
<td>Terminal Objective</td>
</tr>
<tr>
<td>Body</td>
<td>Upon the successful completion of...</td>
</tr>
<tr>
<td>EO</td>
<td>Enabling Objectives</td>
</tr>
<tr>
<td>HeadingEO</td>
<td>Describe the concerns and consid...</td>
</tr>
<tr>
<td>Numbered1</td>
<td>Describe the methods for packagi...</td>
</tr>
<tr>
<td>Numbered1</td>
<td>Demonstrate the ability to use w...</td>
</tr>
</tbody>
</table>

This eliminates the possibility of a document converting with a "NoName" root element.
However, this also limits the solution by restricting the ability to create documents with various highest level elements.

In a paragraph, tags can be assigned based on complex context rules. The converted content is pasted into a blank document that has an EDD imported for this purpose. When the structure is stripped from this document using the Special...Remove Structure from Flow option in FrameMaker, the new paragraph tags are applied. These new, context-sensitive paragraph tags enable much more sophisticated content conversion to be accomplished. Without knowing the intended context of an element, in many cases it is impossible to convert a document and obtain the expected result. If the content is converted incorrectly, the technical writer must know what structure is expected and how to fix the existing content. This can become very costly in terms of required training. Therefore, correctly identifying context became an important aspect of the TEEX implementation.

The content is then structured using a second conversion table that produces structure that resembles what is expected of the final document. However, content conversion alone is not sufficient to produce content that conforms to the production DTD; additional transformation is needed to produce the expected content.
FrameMaker structured applications

Within FrameMaker, a structured application provides several tools for manipulating and validating structured content:

- DTD
- Template
- Stylesheets (pre and/or post process)
- Read/write rules

TEEX utilized three structured applications to successfully transform content into the expected output. The first two applications have very relaxed requirements and are used to alter content so that it will seamlessly import into the production template. FrameMaker contains enterprise functionality that allows this transformation to be accomplished without requiring the purchase of additional software.

Content transformation

The content’s structured application is set to be the first structured application that will save the content in a very flexible XML format. On save, read/write rules are applied and an XSLT stylesheet transforms the content. With this XSLT, it was possible for TEEX to manipulate the content and overcome some limitations that are present because of inherent differences in the way that the original unstructured document format and FrameMaker expose and store information about the content:

- Contents of particular elements are converted into Table Title. This allows a style in the unstructured document that precedes a table to become the Table Title and reduces the number of tasks the technical writer must perform in structured FrameMaker to complete the module.
- Change the root element of the document based on the presence of other elements in the document. Specific styles were created for the document title that identifies the type of content to create. Through XSLT, those specialized titles can be converted into a generic title, while also changing the root element of the entire document. This allowed TEEX to maintain one unstructured template, rather than an instance for each type of content that can be created.
- Convert elements into attribute values.
- Insert sibling elements within the document structure.

Also, some assumptions were made and implemented using XSLT that simplifies many tasks for the end user. These assumptions are easy to change if they are incorrect in a particular instance. For example, tables are one example of a structure that does not easily convert from Word into structured FrameMaker; insufficient information exists in the Word document to identify items as header and footer rows, and column widths and straddled cells are not easily identified in any automatic fashion.

Therefore, TEEX implemented the following functionality for tables using XSLT:

- Every table comes in as the default width of the text frame. This ensures that the table does not exceed the printable width on the page. Also, the columns have equal widths across that table width. Attributes were added so that manual resizing of columns within FrameMaker will roundtrip. This process is transparent to the user.
- If a table contains at least two rows, the first row is converted into a header row. All rows in a Word table come into FrameMaker as body rows because Word does not differentiate between header/ footer rows and body rows. If a header row is not desired, a simple click-and-drag operation in the structure view will convert the row into a body row. Had all rows been brought in as body rows, the process to create a header row is more involved than the solution implemented by TEEX.
- If the table contains at least three rows, the last row is converted into a footer row. As mentioned in the previous discussion of header rows, it is easier to convert a footer row back into a body row than it would be to create a footer row from scratch.
- An element is automatically inserted that serves as a table continuation variable. If a table breaks across multiple pages, the Table Title is repeated at the start of the new page, followed by “(Continued)”. Adding this functionality is an example of how TEEX simplified the solution for the end user.

"The more I work with the template, the more that I realize that this revised process was a really good idea, and well executed."

Kevin Gumienny
Curriculum coordinator, TEEX
Once the stylesheet is run against the document, the content is saved as XML (albeit not XML that conforms to the strict DTD of print-ready materials). This XML document is opened using the second structured application. This allows another set of read/write rules to be applied (other XSLT as well if necessary). Once this additional transformation has taken place, the content, which is valid against a very loose DTD, is then pasted into an instance of the production template that has a much more locked-down DTD. By handling the conversion of the content in this way, the content will convert and be successfully brought into FrameMaker, even if some of the content is not valid against the production DTD. If the initial content had been brought in using the production requirements, any number of potential problems would cause an error that would result in the content processing failing. This would become very frustrating and time consuming for the technical writer because they would have to repeatedly fix the unstructured document and try to convert the content again. This is even more difficult if the solution is not automated because the technical writer is forced to start again at the beginning of the process, saving the document as MIF and then following the other steps of the import process.

Any items that do not conform to the DTD are visible in structured FrameMaker. By using the validation ability within structured FrameMaker, it is possible to quickly identify any missing items and correct the document so that it is valid. The content can then be saved as XML that is valid against the production DTD.
Structured FrameMaker is then used to add advanced functionality to the document: conditional text, complex mathematical equations using the equation editor, table manipulation, assembly of book, etc.

**Automation**

As detailed in the previous section, moving from an unstructured, plain text document to a structured XML document involves a large number of steps that are not necessarily straightforward to the end user. This would create a very steep learning curve and greatly increased training costs. Because TEEX employs over 60 curriculum developers, training and support costs would be prohibitive. Instead TEEX automated these steps using a scripting tool as well as tools and technologies built into Adobe FrameMaker. This automation produced a solution that eased adoption and decreased long term costs. By automating this solution, much of the complexity is hidden from the end users.

**Training**

TEEX developed a training course using the process outlined here to train internal users of the system. This effort produced a structured FrameMaker book document that includes various content modules. TEEX plans to use other Adobe products in the near future to greatly increase the utility of these training materials. The existing FrameMaker book file will be linked into an Adobe RoboHelp® project. Videos will be created using Adobe Captivate® to illustrate concepts covered in the training. Once the RoboHelp project is completed, the solution will be exported as an Adobe AIR® app that can be pushed to each user’s local machine. This will provide interactive and easily accessible training materials that become practical reference material.

Also, utilizing the Adobe AIR format will allow TEEX to easily push updates to end users when the main application is updated. The user will be notified by the application that an update exists and will be able to update immediately to the most current support materials available.

**The future: content management**

In the near future, TEEX plans to implement an enterprise content management system (CMS) to manage and protect its valuable intellectual property. TEEX will also benefit from the sophisticated workflow capabilities, content reuse, and link management that come along with a CMS.

**About Tom Aldous**

www.aldousxml.com

Tom Aldous is an industry-recognized consultant who provides custom services developing DITA and XML solutions with a variety of tools. His specialty is working with structured FrameMaker.