

XML and the Localization Process

By Dan Dube

There has been much publicity in recent months regarding developments in the eXtensible Markup Language (XML) area, and the revolutionary changes that this standard will be responsible for in virtually every industry. Much has been written to describe the impact of XML with regards to e-commerce and content delivery. However, XML is also ideally suited to provide tremendous benefits to companies that have a need to manage and produce content and/or publications in multiple languages. This article will describe how some companies have used existing technology based on XML (and its predecessor, SGML) to realize impressive results, and will provide some vision as to where XML-based technology is heading in the next 24 months with regards to increasing the efficiency of the localization process.

Typical Localization Process

Today, most companies with a need to deliver information to a global audience still rely on a manual, paper-based process for localizing their content. Human intervention is required at all stages, from the origination of source language data through review, quality assurance, and final delivery of localized information. In most cases, authors are working with tools such as Microsoft Word or Adobe FrameMaker and sending complete document instances

to a translation or localization vendor. Because these files are stored and managed at the document level, authors have no easy way to identify only the pieces of content that have changed between revision cycles. As a result, authors continue to send entire documents to the translation vendor for localization, even if only a small portion of the content has been changed.

There are inherent inefficiencies to this process, which typically lead to the following results:

- Longer translation turnaround time
- Higher localization costs
- Lost revenue opportunities

SGML/XML and Document Management Technology Have Helped

Over the last years, some forward-thinking companies have made a strategic investment to create and manage their data in a structured, neutral data format: SGML (Standard Generalized Markup Language) and, more recently, XML. Through a combination of standard products and extensive customization, these companies have created production environments with these characteristics:

- Author and manage information in "chunks" as information objects in a database (rather than entire documents), which enabled the reuse of

information across multiple documents (e.g., a "warning message" can be written once and shared by many documents).

- Write the content once and produce multiple deliverables (e.g., paper, Web, and CD-ROM).
- Track only the information fragments that change between revision cycles and their associated target markets, and only send out changed objects for translation to a specific language/market.

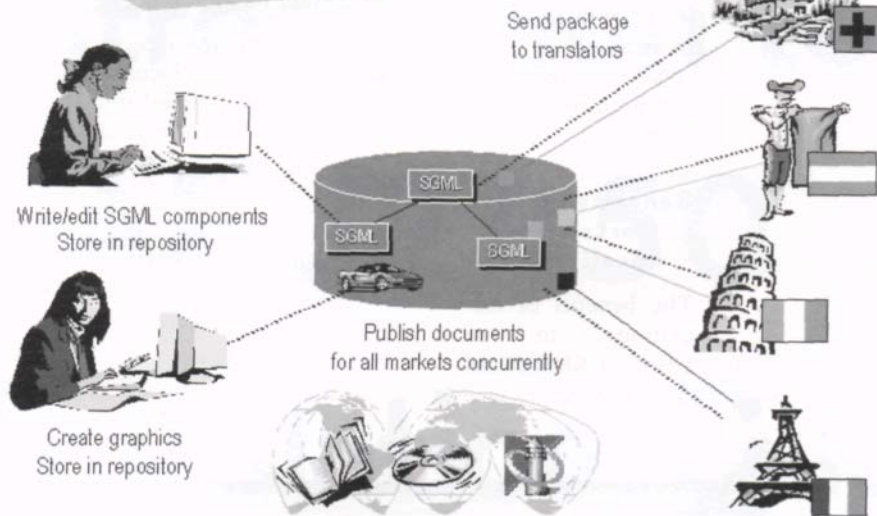
The diagram on the following page depicts the work-flow of this type of environment.

Consider the following statistics from companies that have successfully put an environment like this into production:

- Cummins Engine realized a dramatic 70% reduction in translation costs.
- Tweddle Litho Company, in a project to produce owners literature for a major American automotive manufacturer, was able to cut the time to translate the manuals into 30 languages from six months down to two weeks.

Customized SGML Environment

Store localized data back in repository



SGML/XML and Document Management: Still Room for Improvement

There are solutions available from traditional SGML/XML systems vendors that attempt to replicate the environment depicted in the diagram above. The most well-known of these solutions are Lingua (produced by Chrystal Software as an add-on to their Astoria content management system) and Parlance Ambassador (produced by XyEnterprise as an adjunct to their Content@ product). These systems are sold as a "toolkit", requiring a purchase of base product technology and significant customization services to model the solution to the specifics of your production environment. While these solutions certainly add efficiencies to the localization process, there are still some inherent limitations:

- *These products only support content that is marked up in XML or SGML.*

There is no advertised capability to support legacy data or unstructured data in other formats, such as FrameMaker, Interleaf, or Word.

- *These products only support XML/SGML content that is stored or managed within their repository.*

For example, Chrystal's Lingua solution will only work with structured content stored in the Astoria repository. It will not work with FrameMaker files stored in Chrystal's Canterbury repository system.

- *These solutions can potentially lock a customer into a proprietary vendor solution.*

This is possibly the most disturbing issue associated with these solutions. Even though they are based on the open standards of SGML and XML, it is very difficult to migrate information from one of these repositories to another system in the event you ever want to upgrade to another product in the future. While the core data may be easily exported, it is often very difficult to migrate information about links, metadata, workflow, and version/history information.

- *It is difficult to integrate associated applications and data sets with these products.*

For example, it may be desirable to connect these repositories to your translation memory tool, terminology database, or a translation Web portal maintained by your localization service provider. This would most likely be an expensive customization to these technologies.

The Next Wave: XML Portals

These problems will be addressed by a new generation of open tools, based on the concept of XML portals. In today's world, the localization process involves collaboration of information stored in many related, yet disparate, applications, including:

- Content management systems (e.g., Documentum, Broadvision)
- Translation memory tools (e.g., Trados, STAR Transit, SDLX)
- Machine translation tools (e.g., LOGOS)
- Terminology databases and glossaries
- Digital Asset Management (DAM) systems

XML portal technology shows the promise of being able to provide content owners with the ability to link information stored in these distributed, heterogeneous environments. A portal will simply act as a client to retrieve resources and provide a virtual view of a collection of information. The goals of this technology will be to:

- Connect information collections
- Layer new business rules over existing applications
- Streamline the localization process with automated operations

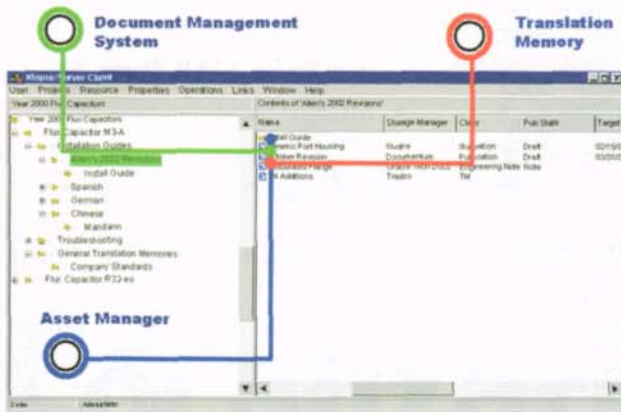
Connect Information Collections

Such tools will enable a user to create a customized view of information that is relevant for his/her work, regardless of where the information may physically reside. The following diagram shows a localization example of this paradigm:

In this example, a localization project manager is provided with a view of information that is relevant for a current translation project. While the user interface may give the appearance of a normal Windows Explorer directory view of information in a folder, the actual information resources physically reside in different applications: a document resides in a content management system, an illustration is stored in an asset manager database, and a translation memory is managed by a TM processor.

Layer New Business Rules Over Existing Applications

With this type of view, it is possible to create links between disparate objects and establish important relationships. In our example illustrated below, we have associated an Installation Guide with a specific translation memory. When the source language version of the Installation Guide file is updated, it will be quite easy to retrieve the appropriate translation memory that is



linked to it. The following diagram depicts this scenario:

Streamline the Localization Process with XML Portals

Once you have linked relevant information objects and created associations between them with appropriate business rules, it is possible to further automate the localization process by assigning operations to linked information collections. The following diagram illustrates this concept:

In this example, a business rule has been established to do the following:

1. Once the "status" of an installation guide has been updated from Draft to Approved, generate a new translation package.
2. Automatically find the translation memory that is associated with the installation guide and add it to the translation package.

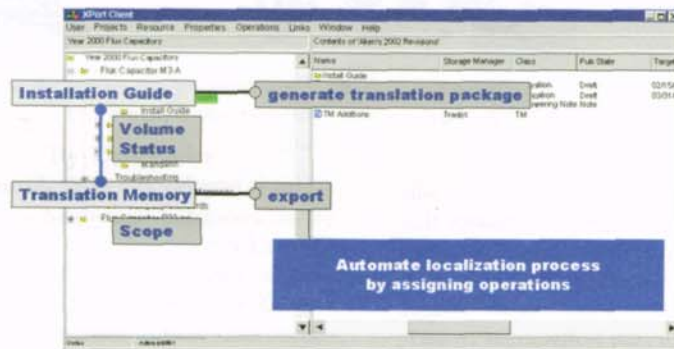
3. Export the translation package, containing both the installation guide and the translation memory, to the translation vendor's workflow system to be localized.

Benefits of XML Portals for Localization

The benefits of migrating to a localization solution based on XML portal technology include the following:

collaboration between content creators and translation vendors.

- Extend the life of existing technology infrastructure by adding new functionality and automation to current business processes.
- Personalize the features of existing technologies (e.g., document management systems, translation memory tools) to the needs of specific users, such as localization project managers.
- Allow for easy migration to new technologies, since the core infrastructure will be based on an accepted international Web standard (XML).



- Maximize reusability and repurposing of information, which will lead to lower translation costs and faster turnaround time.
- Leverage the ability of the Internet to increase efficiency of remote

Conclusions

By now we hope the message is clear: the benefits of XML extend far beyond the e-commerce scenario. The true promise of XML for localization professionals is the ability to streamline and automate the translation process, enable collaboration, and to maximize the capabilities of document management systems and computer-aided translation tools.

Dan Dube is President and COO of Lighthouse Solutions, Inc. He can be contacted at dan@lighthouse-solutions.com

This is an edited version of the article previously published in the LISA Newsletter in September 2000.

