Optimizing the Publishing Workflow

Integration of XML-Aware CMS

By Vasu Ranganathan, Partner and **Shailesh Shah,** Architect, ArborSys Group LLC

The publishing industry is challenged today to create products that are not only highly customized, but are also produced for an ever-growing assortment of formats, all while reducing costs and improving time to market. While new developments in automation and optimization technologies are available to improve the product development workflow, integrating these technologies can be complex.

Workflow Integration Points

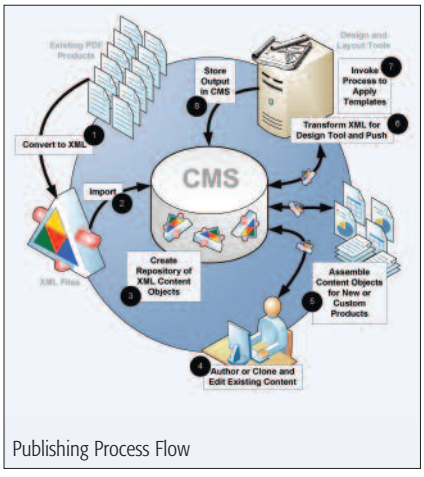
With the incorporation of applicable technologies throughout the publishing workflow, productivity and efficiency can be increased. Integration points include:

- A repository of existing assets, called a data bank;
- An improved editorial process to reuse, create/edit and dynamically assemble customized or new products;
- Intuitive previewing of pages or products for review and approval of assembled content;
- Design and layout tools that automate transformation to press-ready templates; and
- Integration and delivery of content to dynamic delivery engines, such as learning management systems (LMS).

Central to automating the publishing workflow is the integration of an object-oriented, XML-aware content management system (CMS). The primary functions of the CMS are:

- Management of the metadata-tagged chunks of content (or other assets such as images) used by the publishing cycle;
- An XML authoring interface working in tandem with the CMS to allow for the reuse, authoring, and assembly of smaller content chunks, such as lessons, into larger aggregate products, such as chapters and books; and
- Support for content reuse to improve productivity, decrease the duration of the publication cycle, and improve the accuracy and consistency of the content.

The introduction of a CMS to any publishing workflow first requires the development of a well-constructed XML schema expressing the known content components and their interrelationships. The content from existing assets must be mapped to the XML schema and "chunked" into XML components which are then imported into and managed by the CMS. These "chunks" are then used by editors to dynamically assemble new or customized products. Industry-standard XML DTDs or schemas can be leveraged as good starting points and extended where needed.



Publishing Process Flow

CMS technologies such as EMC's Documentum and XML editors such as PTC's EPIC support the functions outlined above.

While establishing a content management system can help create more effective processes, the tool's usability is a critical issue to consider. Typically, users are encountering the XML/CMS conceptual model for the first time. Editors, designers and other users will need training that covers both the introduction of relevant concepts as well as the practical use of the tool. As the tool's graphic interface is designed, keep in mind that users work best with an interface that is intuitive. A "what-you-see-is-what-you-get" (WYSIWYG) approach is most effective.

Additionally, organizations must have a good understanding of their content and how best to "chunk" the content to optimize reuse. Leveraging existing industry XML schema standards developed by consortiums such as W3C can provide a great advantage.

Content Conversion–Creation of a Data Bank

Content reusability begins with a data bank of content assets created from current products. For example, current products may be converted according to the defined XML schema. The resulting XML files may then be imported into the CMS repository while rules automatically assign metadata tags. The content conversion and tagging process as described above can be almost completely automated, but where data is missing or ambiguous some level of manual inference by editors or other subject matter experts may be required.

All parts of the publishing workflow may be made more efficient with the incorporation of a CMS, including presentation design and delivery processes. A CMS can be integrated with industry-standard layout and design tools, such as Adobe InDesign, and with delivery tools such as online or e-learning. These tools working together with the CMS can support designers, and/or be set up to automatically generate final press-ready products by means of pre-configured presentation templates. The final product is then stored in the CMS repository for the purposes of workflow: review, approval and delivery to production. Simultaneously, it becomes a candidate for reuse.

Bringing together 1. an XML-based CMS, 2. an XML authoring tool, 3. design tools for press-ready products, and 4. a dynamic delivery engine for on-line products, is a sound approach to reducing costs and improving productivity. A CMS will automate workflows, speed up production and create abundant opportunities for content reuse. However, this strategy must take into account its users, who will be adapting to an XML-based authoring paradigm usually for the first time. Training and usability are keys to adoption and acceptance.

Organizations must also have a vigorous and comprehensive understanding of the structure of their content if they are to select or define a successful XML schema—the foundation of any XML CMS effort. Content conversion processes must be developed to ensure that the inference of metadata tags is automated as much as possible to minimize manual tagging and qualification by editors.

The incorporation of XML-aware content management systems will bring the publishing industry closer to meeting and exceeding the newest demands of the market: multiple formats, separation of content from presentation, and personalized run-time services for on-line delivery. ▮