

BY BOB DOYLE, Member

DITA Tools from a to Z

Writing in *Intercom's* special issue on content management (March 2007), Bob Boiko asserts, “the knowledge of single sourcing, the knowledge of structured authoring and delivery, [and the] terminology and wisdom of DITA remain largely unknown and unused outside of technical communications departments.”

Boiko called for technical communicators to reinvent themselves and compete on the basis of their value in the global economy. DITA's promise of topic-based, structured authoring is not merely better documentation. It is the creation of mission-critical information for your organization, written with a deep understanding of your most important audiences, which can be repurposed to multiple delivery channels and localized for multilingual global markets. Boiko wants you to “transform what you do from documentation to delivering information products that drive your organization forward.”

To accomplish this goal, you must understand the latest tools in structured writing that are revolutionizing corporate information systems—today in documentation but tomorrow throughout the enterprise, from external marketing to internal human resources. Whether you are trying to push a new product into a new market or are “onboarding” a new employee, the need for high-quality information to educate the customer or train the new salesperson is a challenge for technical writers. You need to think outside the docs!

What is DITA?

The key idea behind Darwin Information Typing Architecture is to create content in small chunks called topics. A topic is the right size when it can stand alone as meaningful information. Topics are then assembled into documents using maps, which are hierarchical lists of pointers or links to topics. The map documents are called *ditamaps*; the pointers are called *topicrefs* (topic references).

Think of documents as assembled from single-source component parts. Assembly can be conditional, dependent on properties or metadata tags you attach to a topic. For example, the *audience* property might be “beginner” or “advanced.”

At a finer level of granularity, individual elements of a topic can also be assigned property tags for conditional assembly. More important, a topic element can be assigned a unique ID that allows it to be reused in other topics. Any other topic can include a *conref* (content reference), and the reusable content element or component will be included at that point.

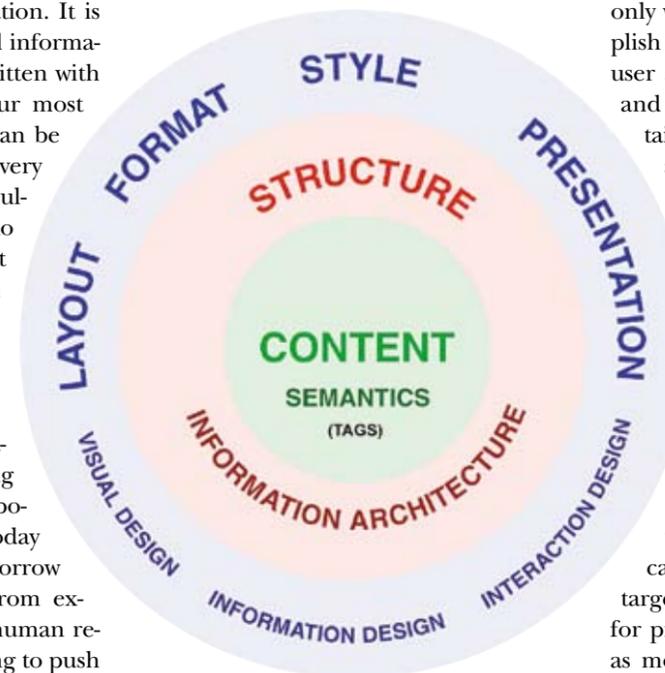


Figure 1. The three levels of XML: the core content with semantic tags (XML), the structure or content model (XSD or DTD), and the styling or presentation layer (XSLT).

DITA is a leading technology for the new “component content management,” which offers dynamic content accessibility at any level of granularity. Components can be as small as a single DITA topic element.

The second basic idea of DITA is that topics can be specialized into a few information types, notably a *task* type and two supporting types called *concept* and *reference*. The task topic is designed to implement a revolutionary idea in mod-

ern documentation called minimalism. Documentation no longer comes as a monolithic book with all the knowledge captured in a linear narrative from introduction to conclusion. It is now conceived as small chunks that answer a specific question, such as “How do I...?” or “What is...?”

A DITA task answers the “how” question with a set of steps and results that walk you through a task. A DITA concept explains the processes needed to understand the steps in a task. A DITA reference collects detailed facts needed for the task.

Minimalist documentation gives users only what they need to know to accomplish a specific task. It assumes that the user already has the basic information and is smart enough to find further details if necessary. Minimalism is ideally matched to today's online user assistance systems, as well as to today's users, who want answers quickly.

Information typing and minimalism are perfectly suited to keeping the corporate message clear, concise, and consistent. In today's global marketplace, reuse of typed topics also means that localization costs are greatly reduced. Finally, typed topics with conditional assembly can be delivered to multiple output targets—HTML for Web delivery, PDF for print, online help, and others such as mobile devices and voice systems—and in multiple languages.

DITA Tools

DITA tools include authoring tools (editors), management tools (content and translation management systems), and publishing engines that fully support DITA. The best DITA tools for technical communicators implement the DITA standard while hiding all the complexity of the underlying XML.

XML requires sophisticated DTDs (document type definitions) or schemas that define all the allowed elements in a content model or information model. Information in XML is “semantically tagged”—that is, elements are tagged according to what they mean, rather

Figure 5. oXygen. This graphic shows the full XML editing view. oXygen can of course show a WYSIWYG view as in Arbortext, but for more tech-savvy editors the full XML file is editable (as in Arbortext, DITA Storm, and XMetaL). Note the attribute inspector pane.

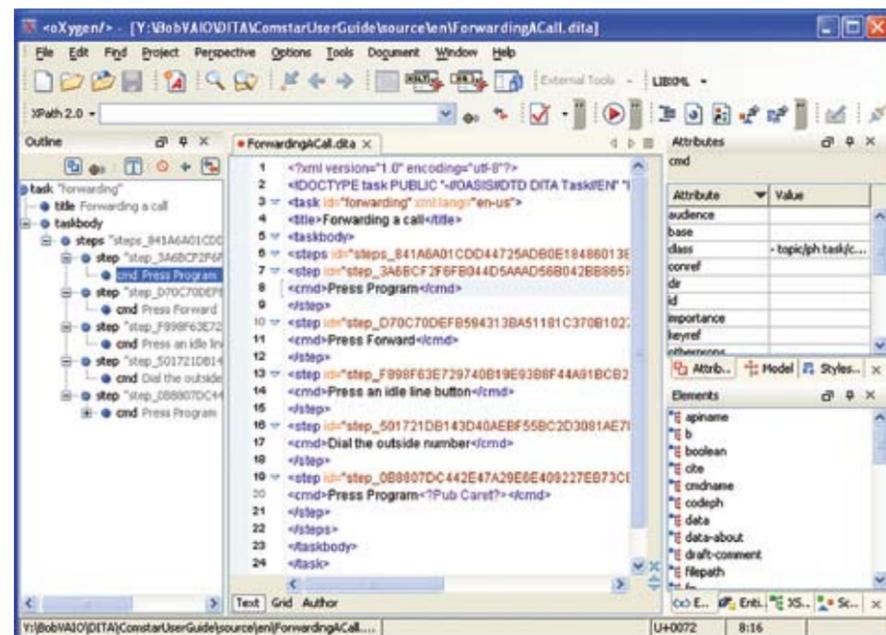


Figure 6. DITA Storm. Note the helpful prompts for missing short description, info, and step results, and for overall task result. An explicit path to the active element is along the top.



formatting. *FrameMaker 8* shows that Adobe is serious about maintaining their twenty-one-year-old desktop publishing software alongside their *InDesign* replacement for *PageMaker*.

JustSystems *XMetaL Author Enterprise 5* (www.xmetal.com). *XMetaL* (see Figure 3) was the first to integrate the DITA Open Toolkit and provide a static publishing solution. The tool quickly cap-

tured significant market share among technical writers moving to DITA. Third parties developed *FrameMaker* connectors to provide high-quality PDF outputs. *XMetaL* now offers RenderX XEP publishing engine integration. *XMetaL* integrates with publishing engines in almost all the leading XML content management systems.

For years, PTC *Arbortext* (formerly Epic) *Editor* (www.ptc.com), like *FrameMaker*, has been a leading tool for structured writing, from SGML and DocBook to XML and DITA. It has been supported by all the major content management systems, not only those that do DITA. Many CMS integrations require the writer to move back and forth between editor and CMS interfaces. *Arbortext Editor's* (see Figure 4) connection to *Arbortext Content Manager*, the industrial-strength PTC CMS (based on the manufacturing CMS called *Windchill*), and the *Arbortext Publishing Engine* offers a seamless authoring and dynamic content publishing experience.

SyncRO Soft *oXygen/*—its name looks like an XML tag—is a powerful XML editor that now has excellent DITA support, including built-in DITA OT (see Figure 5).

Syntax *Serna* (www.syntax.com) is a very affordable multiplatform editor with an excellent WYSIWYG in-context look at the content. It generates a browser view using your XSLT transformations to closely match the final appearance.

XMLmind *XML Editor* (www.xmlmind.com), or *XXE*, is a multiplatform editor offered free for personal use, excellent for those getting started with DITA.

Web/browser-based DITA Editors

DITA Storm (www.ditastorm.com) is a low-cost editor, entirely implemented in Javascript. It works on any operating system running *Internet Explorer* or *Firefox*. It uses an intelligent XSL processor to present the editable document view. It has a ditamap editor, an outline view, and raw XML editing.

JustSystems *XMAX* (www.xmetal.com) is a DOT.net-based ActiveX control that has been integrated into DITA content management systems (Bluestream,

Table 1. DITA Editors

	Adobe FrameMaker 8	Information Mapping Content Mapper	Inmedius DITA Storm	In.vision DITA Studio	Justsystems XMetaL Author Enterprise 5.1	PTC Arbortext 5.3	SyncRO Soft oXygen/ 9.1	Syntax Serna 3.5	XMLmind XML Editor 3.6
Validation	RT,W,OD	RT	RT	RT,W,OD	RT,W,OD	RT,W,OD	RT,OD	RT,OD	RT
Elements in context	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Tags-on view	Yes	No	No	No	Yes	Yes	Yes	Yes	No
Structure/ tree view	Yes	Yes	No	Word Doc Map	Yes	Yes	Yes	Yes	Yes
Visual DITA map editing	Yes	Yes	No	No	Yes	Yes			Yes
Resolved document	No	Yes	No	Yes	Preview	Editable	Preview?	Preview	Preview
Drag and drop	In structure view only	No	No	No	In document view only	Excellent	Yes	Yes	In DITA map only
"Smart" insert	No	No	No	No	Yes (on insert, drag-drop, enter key)	Yes (insert element only)	Yes	No	Yes
Reusable component management and automation	No	No	No	No	Yes	Yes (needs CMS)	No	No	Macros
Spell check	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
Multilingual (Unicode)	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
DITA OT	Yes	Yes	No	No	Yes	No	Yes	Yes	Yes
Platform	Windows, Solaris	Windows	Browser	Windows	Windows	Windows, Solaris	All (Java)	All (C++)	All (Java)
Price	\$899, Education \$319	Standalone \$305, Enterprise \$600	\$300.00	Xpress Author \$300, Studio \$800	\$1,195	\$695	Enterprise \$275, Professional \$225, Academic \$48	Enterprise \$395, Professional \$199, Personal \$89	Professional \$250, Personal free

RT=real-time, W=warning, OD=on-demand, \$=USD

IXIASOFT). *XMAX* users can drag and drop reusable content into a document, drop in multimedia elements such as *Flash* or image files, and conduct collaborative reviewing and approvals.

Word-to-DITA Editors

Information Mapping *ContentMapper* (www.infomap.com). Building on forty years' experience training tech writers to structure their writing, Informa-

tion Mapping has customized Microsoft *Word* to constrain content to valid DITA "under the hood." Drop-down menus become context-sensitive, guiding the writer to create DITA topics using the industry's most popular content-creation tool.

In.vision *DITA Studio* (www.invisionresearch.com) builds on the DITA OT to publish documents edited with their *Xpress Author for Word*, a customization

of Microsoft *Word* that constrains content to valid DITA. *DITA Studio* is integrated with Documentum, Open Text, Astoria, Vasont, Mark Logic, Trisoft, and SharePoint.

DITA Content Management Systems

Following is a brief list of the major content management systems that provide specific support for DITA publishing. As a technical communicator, you

Table 2. DITA Content Management Systems

	Native XML database	SaaS*	Estimated Cost	Integration					
				Arbortext	FrameMaker	XMetaL	DITA OT	Idiom	Trados
Astoria On Demand	Yes	Yes	\$12K + up	Yes	Yes	Yes	Yes	Yes	Yes
Author-it	No	No	\$5K – \$500K	[1]	[1]	[1]	No	Yes	Yes
Bluestream XDocs	Yes	Option	\$5K + up	Yes	Yes	Yes	Yes	No	No
DITA Exchange	Yes	Yes	\$500/mo + up	[2]	[2]	[2]	Yes	No	Yes
DocZone	Yes	Yes	\$7K + up	Yes	Yes	Yes	No	Yes	Yes
Inmedius Horizon	Yes	No		Yes	Yes	No	No	No	No
IXIASOFT DITA CMS Framework	Yes	No	\$65K + up	No	No	Yes	Yes	No	No
PTC Arbortext Content Manager	No	No	\$100K + up	Yes	No	No	No	No	No
SiberLogic SiberSafe	Yes	Option	\$250/mo. + up	Yes	Yes	Yes	Yes	Yes	Yes
Trisoft Infoshare	Yes	Option	50K€ – 250K€	Yes	Yes	Yes	Yes	Yes	Yes
Vasont	No	Option	[3]	Yes	Yes	Yes	Yes	Yes	Yes
X-Hive Docato	Yes	Option	\$75K + up	Yes	Yes	Yes	Yes	Yes	Yes
XyEnterprise Content@	Yes	No	[3]	Yes	Yes	Yes	Yes	Yes	Yes
*software as a service \$=USD, €=Euro	[1] Author-it has a built-in editor.			[2] Has API for XML-editor integration.					[3] Company will not disclose.

should be familiar with most of these since they are likely to be the way your DITA-structured content gets processed and published in multiple formats. DITA content management systems range in cost from several thousand dollars to a few hundred thousand dollars. Many can now be rented on a monthly basis, with no minimum contracts, for as little as a few hundred dollars per month. (Software rental is known as SaaS, or software-as-a-service.)

Astoria On Demand (www.astoriasoftware.com) builds, manages, and assembles dynamic, globalized DITA content from top DITA editors such as *Arbortext*, which have Internet access to the Astoria CMS. Astoria is SaaS.

AuthorIT (www.authorit.com) was doing single-source, reusable component content management long before

DITA. Its help system output has made it a major help authoring tool. Its DITA support is idiosyncratic and controversial, but *AuthorIT* is widely used by tech writers, who will find it easy to move to DITA with *AuthorIT*'s familiar tool.

Bluestream XDocs (www.bluestream.com) is a low-cost CMS targeted at budget-conscious organizations. It offers integration for several editors, the DITA OT, and a native XML database.

DocZone (www.doczone.com) was the first hosted-only (SaaS) XML and single-source publishing solution. It features an integrated translation management system (TMS) from XML International.

IXIASOFT DITA CMS Framework (www.ixiasoft.com) features a native XML database CMS with *XMetaL*, *<oxygen/>*, and the *XMAX* Web-based editor. Users also benefit from a drag-and-drop dita-

map editor, localization support, and a publishing engine.

Microsoft *SharePoint* (www.microsoft.com) is the underlying CMS for DITA solutions like *Invision DITA Enterprise Suite* (www.invisionresearch.com) and *DITA Exchange* (www.ditaexchange.com).

PTC *Arbortext Content Manager* (www.ptc.com) is a version of PTC *Windchill*, now an integral component of the PTC Dynamic Publishing System, which also includes *Arbortext Editor* and the *Arbortext Publishing Engine*.

SiberLogic SiberSafe (www.siberlogic.com) includes semantic knowledge modeling technology, a novel method for tagging and retrieving content components using an inference engine to identify the meaning of each component.

Trisoft Infoshare (www.trisoft.be) features out-of-the-box full DITA support,

integration with DITA editors and publishing engines, built-in translation management, and a link manager.

Vasont (www.vasont.com) includes a visual user interface that lets editors access Vasont's functionality from the main menu of XML editors, including *Arbortext*, *FrameMaker*, *XMetaL*, and *Word*. Multilingual translation management.

X-Hive *Docato* (www.x-hive.com) offers a DITA Starter Kit preconfigured for DITA schemas. It is the CMS offered as a hosted SaaS by DocZone.

XyEnterprise Contenta (www.xyenterprise.com) was the first reusable component-based content management system. It is used primarily by very large corporate publishing operations.

DITA Publishing Engines

A publishing engine is a special application that takes your DITA topics and ditamaps and serves them out automatically to print and the Web. They can cost from free (the Open Toolkit) to more than one hundred thousand dollars. Fully automated publishing solutions integrated with an XML CMS—such as those from Astoria, Vasont, and XyEnterprise—or integrated editing, styling, publishing, and content management systems such as PTC *Arbortext* can cost millions of dollars when implemented for thousands of users and multiple publishing servers.

Adobe *FrameMaker Server* (www.adobe.com) is a license to use their formatting and PDF production engine in an automated publishing environment. Documentation for the Adobe *Creative Suite* is an example of the high-quality print output.

Arbortext Publishing Engine (www.ptc.com). Known for years as E3 (Epic E-Content Engine) when the *Arbortext* editor was called *Epic*, the *Arbortext Publishing Engine* is a market leader for structured publishing (SGML, then XML, now DITA).

DITA Open Toolkit (dita-ot.sourceforge.net) is a free open-source reference implementation of DITA processing to XHTML, PDF, and a variety of help formats. The OT is integrated into many editors (*FrameMaker*, *<oxygen/>*, *XMetaL*) and content management systems

DITA Resources

The OASIS DITA Focus area (dita.xml.org) is the official community resource site by the maintainers of the DITA standard.

DITA Infocenter (www.ditainfocenter.com) has language elements, architectural specs, and the OT User Guide in a searchable help format.

DITA Users (www.ditauers.org) and its network of DITA support sites offer resources, a newsletter, a blog, a wiki, and tutorials on DITA. Members get an online workspace folder with sample docsets, Web authoring, and publishing to HTML, PDF, and Help.

Community mailing lists include dita-users@yahoo.com and framemaker-dita@yahoo.com, and the single-source discussion list (stc-single-sourcing-l) at STC.

(Astoria, Bluestream, IXIASOFT, XyEnterprise). You can download the OT and install it for free on your computer, to get started with topic-based writing and publishing. Or you can use an online (SaaS) version by joining DITA Users (www.ditauers.org).

Elkera *XML Print* (www.elkera.com) simplifies style sheet development and maintenance compared to traditional approaches using predeveloped style rules for DITA. Nontechnical users can adjust page layouts and styles for their DITA style rules by editing a template document in *Word*.

XyEnterprise *XML Professional Publisher* (www.xyenterprise.com), or *XPP*, has out-of-the-box support for DITA. *XPP* automated publishing technology was the first to add fully interactive WYSIWYG editing.

DITA Localization and Translation Management Systems

Idiom Technologies *World Server Globalization Management System* (www.idiominc.com). Idiom pioneered DITA multilingual publishing, first with *Au-*

odesk and then for Adobe's conversion of their *Creative Suite* documentation to DITA. Today they focus on their *World Server*, integrated alongside DITA in many CMS's.

The Future of Publishing

Technical writers are typically good writers but poor techs, and IBM's free DITA Open Toolkit is easy to install only for programmers. Further, installing the OT on a laptop or desktop limits its use to one individual. Many writers can share an OT on a Web server, and their publishing deliverables can be seen immediately on the Web. SaaS is the model for highly scalable content publishing in the future. 

SUGGESTED READINGS

Ethier, Kay. *XML and Framemaker 8*. Apress: 2008.

Hackos, JoAnn. *Information Development: Managing Your Documentation Projects, Portfolio, and People*. Wiley: 2006.

Hackos, JoAnn. *Introduction to DITA, Arbortext Edition*. Comtech Services: 2007.

Hanna, Robert. *DITA Pocket Guide*. *SiberLogic*: 2006.

Hargis, Gretchen, et al. *Developing Quality Technological Information: A Handbook for Writers and Editors*. 2nd Edition. IBM Press: 2004.

Linton, Jennifer, and Kylene Bruski. *Introduction to DITA*. Comtech Services: 2006.

van Raaphorst, Anna, and Richard H. (Dick) Johnson. *DITA Open Toolkit User Guide*. VR Communications: 2006.

Bob Doyle (bobdoyle@skybuilders.com) is the editor of CMS Review and a contributing editor of EContent Magazine. He is a member of the OASIS DITA editorial board. He founded the CM Professionals organization in 2004 and DITA Users in 2007.

Discuss this article online at stcforum.org/viewtopic.php?id=xxxx