



The XML Strategist

Casting a critical eye on the Next Big Thing in technical publishing.

SARAH S. O'KEEFE, Column Editor

The Hidden Cost of DITA

BY SARAH S. O'KEEFE, Associate Fellow

In the past few years, we have implemented both DITA-based and custom XML solutions for our customers. Given the right set of circumstances, DITA provides an excellent foundation for structured content. But I seem to be in significant disagreement with DITA advocates about how often the “right set of circumstances” is present.

The Gartner Hype Cycle, shown in Figure 1, describes the “common pattern of human response to technology.”

In my opinion, DITA appears to be near the Peak of Inflated Expectations. (For more on the hype cycle, visit Gartner’s Web site: www.gartner.com/DisplayDocument?doc_cd=130115.)

The Peak of Inflated Expectations is the point where technology seems incredibly exciting, but none of its flaws—and all technology has flaws!—are being given serious consideration.

I hope this article will allow us to begin discussing DITA implementation challenges.

Content Modeling

Too many people see the DITA architecture as a shortcut to avoid content modeling. The logic appears to be something like this: “The DITA designers are smart people who designed something very useful that will work great for my content.” I agree that the DITA designers are smart and that DITA is a significant achievement. It’s the last part that concerns me.

Here’s a brief checklist of content modeling questions:

1. Is the DITA content model appropriate for my content?
2. How much work would be required to customize or specialize DITA to make

it match my content exactly?

3. How important is the content model to the information I create?

Most important, I think you need to ask this question:

4. How different would the content model look if I built it from scratch?

One trade-off in choosing DITA (or any other standard) is that you must conform to the general worldview of the standard you are implementing. DITA was initially designed to create topic-oriented, modular content that describes software applications. If your content is currently not topic oriented, a DITA implementation means a significant shift in how your content is written and organized. Although it may be the

right approach, the transition will be expensive and time-consuming and should be included in the cost analysis.

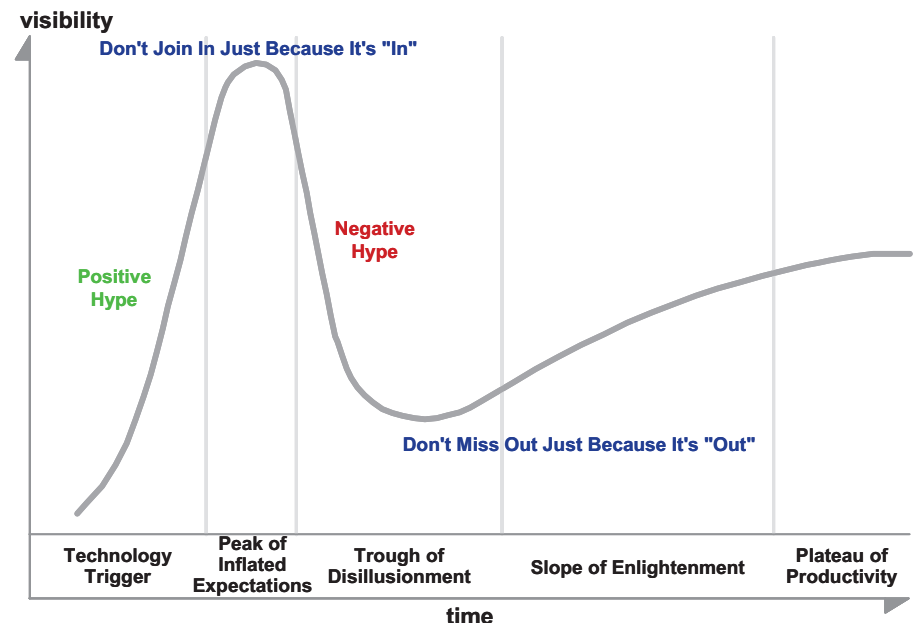
In addition, I believe that there is another important consideration—if you decide to build on DITA without fully understanding your content model, you will end up with a DITA-shaped box that may or may not be the correct shape for your content.

Specialization

DITA specialization lets you customize the DITA structure without breaking the output processing. Because you create new elements that are explicitly based on existing elements, default processing is still available. However, when you use specialization, your new element must be congruent with the parent element. That is, the specialized element must use a structure that is valid for the original element. The content model must either match the original element or be a subset of the original element (for example, you could eliminate optional elements to make the new element stricter). Your specialized element must not use a structure that is invalid for the element from which you specialized.

(Continued on page 40)

Figure 1. Where is DITA on the hype cycle?



Source: Gartner (July 2007)

(Continued from page 35)

Evaluating DITA Features

DITA supports sophisticated reuse and conditional processing techniques. The question is, do you or will you need these features in your workflow? If you need them or think you might need them in the future, then score major points for DITA.

But what if you have minimal reuse and very simple conditional requirements? DITA may be more than you need, and there is a cost associated with maintaining it. The DITA Open Toolkit is especially challenging.

Here are some questions to ask as you evaluate DITA features:

1. Do DITA features provide useful benefits for my publishing workflow?
2. Does the DITA Open Toolkit provide all the output formats I will need? How much customization will be required to make the output meet my requirements? Will I have to create formats, and is there anything related in the Open Toolkit?

Keep in mind that some customizations are easier than others. Changing formatting, such as the appearance of a particular tag, is relatively easy for the HTML outputs. Any customization for PDF output is going to be more challenging than the HTML equivalent.

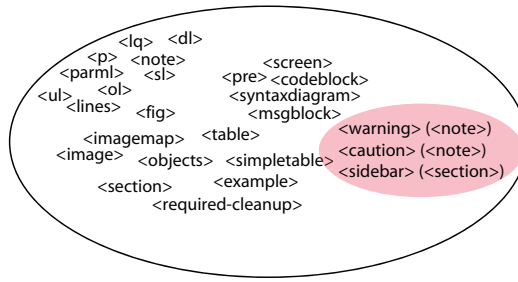
If you need to change link processing, conditional text processing, or map processing, expect a significant effort. You'll need someone who knows XSLT and preferably the specifics of DITA Open Toolkit XSLT. For PDF processing, you'll need someone who can write XSL-FO (Formatting Objects) code, and there aren't very many people out there with those skills.

Configuration Effort

A workflow based on open-source tools obviously cuts your software licensing costs. For some of our customers, the ability to distribute the code onto many servers without worrying about licensing is hugely appealing.

But DITA configuration is challenging, and it requires a significant effort. There are, of course, consultants who

Figure 2. Specialization lets you build on existing elements.



can help with DITA implementation, but they generally don't work for free. If you take on DITA configuration on your own, expect to spend a significant amount of time getting everything to work.

Some questions to consider:

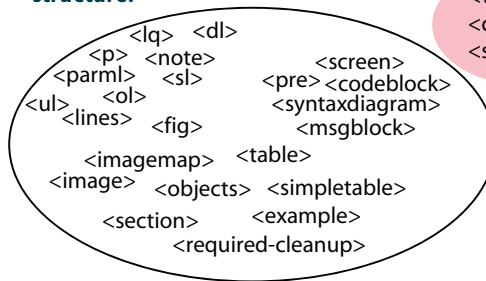
1. How much specialization will be required to match my content model?
2. How much configuration will be required to get the output I need?
3. Who will do the necessary configuration, specialization, and implementation work?
4. Is there a budget—time or money—for this work?

To produce PDF output, you will need a Formatting Objects processor. An open-source processor is available, but you will get higher-quality output from a commercial rendering engine. Even in otherwise strictly open-source workflows, many companies turn to commercial software for PDF output because the output quality is much better than the open-source options.

Free But Not Cheap

DITA is a tremendous technical achievement, and if you are considering implementing XML, you should evaluate whether DITA is an appropriate foundation for your efforts. But

Figure 3. Customization creates elements that are outside the original structure.



XML implementation, with or without DITA, is challenging. One common recommendation is to implement DITA without any specialization, get writers accustomed to working in the new environment, and then use specialization later to refine the content model to fit your content better.

It's my opinion that the content modeling phase is actually the most important part of the entire XML implementation process, and deferring it is a mistake. Instead, begin by understanding your content requirements. Once you have done that, you can evaluate how much work is required to make your content model fit into DITA. You may decide to build with DITA, to use a different standard, or to build your own XML vocabulary. I feel that using DITA as is, without conforming it to your content requirements, is a mistake for all but the most simplistic documentation efforts.

DITA may be free, but it's not cheap.

Sarah O'Keefe is founder and president of Scriptorium Publishing Services, Inc. (www.scriptorium.com), based in Research Triangle Park, North Carolina. The company is focused on implementing tools and processes to optimize publishing work flows. Services include developing and deploying XML-based structured authoring environments, configuring authoring and publishing tools, and providing technical training. Sarah's publications include Publishing Fundamentals: FrameMaker 7, The WebWorks Publisher Cookbook, Technical Writing 101, FrameMaker for Dummies, and numerous white papers (available at www.scriptorium.com/papers.html).

Sarah is an STC associate fellow and member of STC's Carolina, India, TransAlpine, and U.K. Chapters and of the Consulting and Independent Contracting, Europe, and Management communities. You can reach her at xmlstrategist@scriptorium.com.

Author's note: Many thanks to my Scriptorium coworkers Alan Pringle, David Kelly, Karen Brown, Simon Bate, and Sheila Loring for their contributions to this article.

Discuss this article online at stcforum.org/viewforum.php?id=51.