



The PDF Landscape for DITA Content

BY SARAH S. O'KEEFE | *Associate Fellow*

THERE ARE NUMEROUS alternatives for producing PDF output from DITA content. The approach you choose will depend on your output requirements—do you need images floating in text, sidebars, and unique layouts on each page? How often do you republish content? How much content do you publish? Do you need to create variants for different audiences? Do you provide content in multiple languages?

This article describes several common approaches and what requirements they support best. Your options include the following:

- ▶ The DITA Open Toolkit (OT) with an Extensible Stylesheet Formatting Objects (XSL-FO) processor (Apache FOP, Antenna House XSL Formatter, and RenderX XEP)
- ▶ XML authoring tools that work with the DITA OT and an XSL-FO processor to produce PDF files. They let you author XML and create PDF files from an interface (Just Systems XMetaL Author, SyncRO Soft oXygen, and Quark DITA Studio)
- ▶ Conversion tools that produce PDF files from DITA files (WebWorks ePublisher)
- ▶ Help authoring tools that import DITA files and have built-in PDF file conversion capabilities (MadCap Flare and Adobe RoboHelp)



This column casts a critical eye on the Next Big Thing in technical publishing. Please send comments and suggestions to xmlstrategist@scriptorium.com.

- ▶ Page-based publishing software that imports DITA files and has built-in PDF file conversion capabilities (Adobe InDesign and FrameMaker)
- ▶ High-capacity enterprise publishing tools that work with the DITA OT and produce PDF files (SDL XySoft XML Professional Publisher and Arbortext Publishing Engine)

DITA Open Toolkit with XSL-FO Processor

The DITA Open Toolkit includes support for PDF output via XSL-FO. By default, the output created through the Open Toolkit is ugly, and customizing the XSL-FO code is a daunting task. The advantages of the Open Toolkit are automation and licensing cost. You run the Open Toolkit from the command line, and it's possible to integrate the Open Toolkit with automated build systems. If you use the free FOP processor, you can generate PDF without any software licensing costs. The commercial FO processors cost up to \$5,000 but have better functionality than FOP. Configuring the Open Toolkit to produce even reasonably attractive pages requires significant technical skills and is not for the faint of heart.

DITA-capable XML Authoring Tools with PDF File Conversion

DITA authoring tools, such as XMetaL Author and oXygen, provide a way to run the DITA Open Toolkit from within the authoring environment. This approach is friendlier than requiring an author to run a command line to kick off PDF generation, but the configuration process still requires you to modify the DITA Open Toolkit. The authoring tools do provide a way to specify some parameters (such as conditional settings). The software licenses include a rendering engine—FOP for oXygen, the XEP (a more robust commercial engine) for XMetaL.

WebWorks ePublisher 2009

WebWorks ePublisher is a conversion tool that includes the DITA Open Toolkit and XEP, so once again, you face a difficult Open Toolkit configuration project to produce the output you want. If you use ePublisher for other outputs or want some of the automation that ePublisher can provide, this might be a good option.

Another tool in this class is DITA2GO. As this article was written in March 2010, DITA2GO has announced that they are working on PDF output.

Help Authoring Tools with DITA Import and PDF File Conversion

Adobe RoboHelp and MadCap Flare both have the ability to import DITA content and render it to PDF. The process of configuring these tools is much easier than working in the DITA Open Toolkit. However, since these tools are intended primarily for help output rather than PDF, the options for print formatting tend to be somewhat limited. These tools also allow you to generate HTML-based help (often called WebHelp), so if you need that output and are happy with a simple, but respectable, PDF look, a help authoring tool might be the way to go.

If you have DITA specializations, you will probably need to transform your files back to unspecialized DITA before importing into Flare.

Page Layout Software with PDF Output

The traditional page layout tools, including FrameMaker, InDesign, and QuarkXPress, can accept DITA content. Once the information is in the page layout application, it is treated like any other content, so you can take advantage of all the layout features. FrameMaker's DITA support is much better than the other page layout tools and can be further improved with the third-party DITA-FMx plug-in.

The major advantage of page layout software is that you can see the exact layout and pagination and make adjustments before creating the PDF output. This workflow increases the cost of production, but may be worthwhile for highly designed publications.

Enterprise Publishing Tools

For enterprise publishing requirements, consider tools such as XML Professional Publisher (XPP) or Arbortext Publishing Engine. XPP is intended for high-volume, intricately formatted publications, such as in financial publishing, and allows users to make adjustments to formatting before generating the final output. The Arbortext Publishing Engine also has high-end features such as change bars and column-wide footnotes. Enterprise-class tools can address formatting requirements that none of the other options will support.

Conclusion

Evaluate the following factors to select your DITA-to-PDF file process:

AUTOMATION. If automated production is a priority, avoid the page layout tools and the temptation to reach into the intermediate layout files. Instead, consider the DITA Open Toolkit and choose your FO processor based on formatting requirements.

FORMATTING REQUIREMENTS. Are ligatures, attractive justification, and hyphenation critical? Do you have requirements, such as mixed columns on a single page, that the XSL-FO processors cannot support? You probably need page layout software. On the other hand, if your

formatting requirements are simple, you can probably use any of the options discussed here; look at other evaluation factors.

DIFFICULTY OF CONFIGURATION. If you want to minimize the difficulties in formatting your output, consider a help authoring tool. If you want a technical challenge, go with the Open Toolkit.

FORMATTING ADJUSTMENTS. If hand-tweaking the formatting before generating the final output is a requirement, do not use any of the options based on the Open Toolkit. You can adjust formatting in the various help authoring, page layout, and enterprise publishing tools.

COST. The combination of the DITA Open Toolkit and the FOP processor is free. All of the other tools have at least some cost.

EXISTING TEMPLATES. If you already have formatting templates in a specific tool, consider using that tool to produce your DITA PDF output. For example, an organization that already has an unstructured FrameMaker template or an InDesign template that meets all of their requirements might stay in those tools to take advantage of the existing template files.

LANGUAGE SUPPORT. If you need to support a wide variety of languages, verify that your languages are supported or can be supported by the tools you are considering. In particular, right-to-left languages, such as Hebrew and Arabic, are not widely supported. The DITA Open Toolkit actually has excellent language support.

Many technical communicators equate XML or DITA authoring with ugly PDF output. The default output provided through the DITA Open Toolkit is certainly rudimentary. However, there is no technical reason that PDF from the Open Toolkit should be ugly, and XSL-FO consultants are available. If automation is not a high priority, a help authoring or page layout tool could provide a reasonable alternative with a smaller learning curve. ■

This article is a condensed summary of "Creating PDF files from DITA content," published by Scriptorium Publishing and available at www.scriptorium.com/resources/white-papers.

Sarah O'Keefe (xmlstrategist@scriptorium.com) is president of Scriptorium Publishing Services, Inc. (www.scriptorium.com). The company provides expert advice and collaborative development to streamlining publishing processes, resulting in cost reduction and quality improvements. Visit her blog at www.scriptorium.com/blog or follow her on Twitter: sarahokeefe.

