Can you explain that again? DITA for beginners

DITA is a difficult thing to explain to the uninitiated. It is difficult because we expect it to be a product or a technology, when it is actually a standard and a methodology. DITA provides an approach to technical writing that embraces best practice ideals such as modularity, single-sourcing, and content re-use. The reasons for moving to DITA are business-focused.

Photo: Sebastian Duda

By Tony Self

Modular documents are efficient

Wouldn't it be good if you could write documentation in components, and then build those components into different documents depending on requirements? Such a modular document approach can be applied effectively in many different scenarios. The approach is very efficient, because you only have to write and maintain a piece of information once. If you make a change in one component, the change flows through to every document that uses that component. The term re-use is sometimes used when describing this feature. Modular documentation is not a new idea. It was even used before computerization, in the "Typewriter Age", within a writing methodology know as STOP: Sequential Thematic Organisation of Publications. The STOP methodology called the document components "topical units of discourse"; we now usually refer to those components as "topics". Help systems have always been built upon the concept of topics, and as Help Authoring Tools became more sophisticated, modular document features were progressively introduced. The World Wide Web is also built around the concept of topics, and the unrestricted ability to link from one topic to another means that the Web also embraces the idea of modularity. However, in the parallel universe of print-based documentation, modularity has not been as accepted. Document formats such as Microsoft Word's are based on the document being the primary unit, not the topic. DITA is a methodology which includes a document format, and it is designed specifically for modular documents. In other words, DITA makes modularity really simple for all types of document delivery methods, including Web, Help and print-based. DITA has two main types of information structures: topics and maps. Maps are simple specifications for a document, listing the topics that make up the document in the order and hierarchy in which they are to appear.

Information typing means more usable documents

A number of writing methodologies favour the idea of segmenting information based on its nature (and purpose). The underlying theory is that people read manuals to satisfy specific needs. In some cases, they might need to find out how to do something. In other cases, they might need to see how something works. In other cases yet, they might need to look up a code to enter. Rarely will someone open a manual because they want something to read. Satisfying a reader's particular need can be achieved by separating the "how to" information from the "how it works" information as well as from the "facts" information. In the Information Mapping approach developed in the 1960s by Robert Horn, there were a handful of information types, including principle, process, procedure, concept, and structure. Years later, Microsoft was using seven information types in its documentation, comprising conceptual, FAQ, glossary, procedural, reference, troubleshooting, and tutorial. In the DITA approach, there are three base information types: task ("how to"), concept ("how it works"), and reference ("facts"). Perhaps surprisingly, content of most manuals and Help systems fit easily into those three simple categories. However, when those three simple types are not appropriate for the content, DITA allows for the "evolution" of new information types. If you have nothing better to do, you could create new information types yourself, but in most cases, new types are created within industries or areas of interest. Information typing in DITA also guides you towards consistent content that embraces best practice technical writing techniques. This is made possible by the application of rules in a document. For example, when documenting a task, you have to include at least one step. If you don't include a step, the topic won't save! This enforcement of writing rules is in turn made possible by the fact that DITA is an XML-based document format, and XML was designed for this. The term used in XML for enforcing document rules is validation. Producing quality documentation within a DITA approach still relies heavily on your skills as an author; information types and validation make it easier for you to get it right every time.

Single-sourcing through semantic mark-up

The term single-sourcing means different things to different people. Fundamentally, most would agree that it means using the same source content to produce different deliverable products. It is an extension of the idea of modular documents to include different delivery modes; not only can the same content appear in different publications, it can also appear within entirely different media. An instruction might appear within a printed user guide, within a Help topic, on a Web page, and in an ePub. It could appear in the second level of a quick start guide, and in the fifth level of an administrator's guide. For single-sourcing to be simple, content can't be marked up with formatting instructions. It's no good marking a topic title as Heading 2 if it might need to be marked in a Heading 4 in a different publication. Text can't be marked in 12 point if it might end up appearing on a mobile phone screen where 12 point is too large. DITA bypasses this potential roadblock to effective single-sourcing through semantic mark-up. Instead of marking up text based on how it should look, you mark up text based on what sort of text it is. Titles are marked up as titles. Pre-requisites are marked up as pre-requisites. Steps are marked up as steps. Warnings are marked up as warnings. File names are marked up as file names. Semantic mark-up allows the separation of content and form. The form (or style) is added, based on rules that map semantic mark-up with presentational styles, during the publishing process. The publishing process is automated; it is pretty much a one-click process, once the publishing mapping rules for the organization have been created. Want a PDF? Click PDF. Want an ePub? Click ePub. Want Eclipse Help? Click Eclipse Help. One of the challenges for technical communicators is changing focus from form to content. It may sound easy, but it is quite a transition to move from style-based authoring to semantic authoring. The benefits are many. By automating the formatting process, you can spend more time on the words and phrasing, rather than on the fonts, alignment and numbering! This leads to better writing quality, and more consistent presentation.

DITA is community-owned

Are you sold on DITA? It will save you time, help you produce better quality documents, free up more time to spend on writing, and make your professional life easier! So where do you go to buy DITA? This is where we need to have another shift in thinking. DITA

is a standard, not a product. And it's an open source standard. That means that DITA is free: free-as-in-freedom, and free-as-in-beer. Open standards are created and maintained by a community, rather than by a corporation. DITA is "owned", if that's the right word, by the technical writing community. The standard is managed through a not-for-profit standard body called OASIS, and is guided by a group of volunteers on the OASIS DITA Technical Committee. To adopt DITA, we need to find authoring tools that support the DITA standard. Because the DITA standard is open, you can choose from dozens of authoring tools, including FrameMaker, Arbortext, XMetaL, oXygen, Serna, XXE, DITA Storm, Xopus, and many others. You can even switch from editor to editor, mid-topic if you like! But too many choices can be confusing, particularly to the newcomer. And although DITA is free, commercial DITA authoring tools are not. They can vary in price from less than USD100 to more than USD1000. The separation of content and form in DITA has generally led to different types of tools for authoring and publishing. Rather than choose one tool for your DITA workflow, you might need to choose two or three.

Re-use, re-use, re-use

One of my favourite pieces of DITA jargon is WOOO: Write Once and Once Only. Nearly all features in DITA aim to reduce your workload, and one way this is done is by eliminating repetitive work. Once you have written a particular phrase, block or topic (whether that be a product or company name, warning, set of steps, topic or chapter), you should never have to write it ever again. DITA has plenty of mechanisms for content re-use, many with exciting names such as transclusion and indirection. (The unfamiliar terms disguise the fact that these features are clever in their simplicity.) You might have encountered variables in your current authoring environment... you can think of DITA's re-use features as variables on steroids! Any type of DITA content fragment can be re-used. Paragraphs can be re-used, notes can be re-used, phrases can be re-used, terms can be re-used, maps can be re-used, index terms can be re-used, and whole topics can be re-used. This means that the idea of modular documentation can be extended way beyond the simple re-use of topics in different publications, to

re-using anything that would otherwise have to be re-typed or copied. Re-use makes it much easier to keep content up-to-date, because you only have to make any change once.

You might need a CCMS, whatever that is...

Once you embrace the modular document, heavy duty single-sourcing, and re-use approaches that are integral to DITA, managing your content can become a challenge. How do you know if someone else in your team has already written a similar topic? How do you know where the product name variables are stored? How do you know which author in your team wrote a particular topic? How do you know in which publications a particular topic appears. That management challenge can be addressed with a software tool; in this case a Component Content Management System. You may not have seen that extra "C" in front of "CMS" before, but it means a type of CMS that can work with modular document components.

DITA may seem complex

DITA is often said to be complicated, and too complex. The current DITA standard has over 500 semantic elements... how can you be expected to remember what they are all for? DITA is different in many ways to earlier documentation approaches. That difference is a barrier to adoption. To take advantage of DITA, you need to rethink the way you've been approaching documentation. You need to understand the principles of the separation of content and form, and be prepared to let go of the form part! You need to write within supra-organizational standards, and embrace the ideals of open source. You need to let go of a one-tool-fits-all philosophy, and work with a set of tools appropriate to you. You need to learn the purpose of a small number of semantic mark-up elements (nowhere near 500, by the way), and when to apply them. You need to see a documentation project as part of a library, rather than as an individual publication. You need to work at a smaller level of granularity, and understand how that allows re-use to make your life easier. Why, you might even need to learn a bit about XML, but that really depends on what tools you choose and whether XML interests you. Some say that

DITA is restrictive, because it is full of rules and standards and validation; DITA stifles creativity, they say. I think it's almost the opposite. Think about haiku poetry. It's full of rules about syllable weight, phrases and meter. Does anyone ever say that haiku stifles creativity? Like haiku, DITA promotes creativity.

What is DITA?

DITA is a methodology and an open standard, built on XML, and maintained by the technical writing community. It makes it possible to apply technical writing best practices such as modularity, single-sourcing, and content re-use, primarily through the separation of content and form. DITA allows the publishing process to be automated, reducing the author's workload. Although the DITA standard is free, authoring and publishing tools are commercial. You may need to use different tools to work in DITA. It may seem complicated at first, but when the ideas behind it start to click in your mind, it suddenly becomes simpler. Finally, when used as designed, DITA results in better quality writing, at a lower cost. Oh, I forgot to mention one thing. DITA stands for Darwin Information Typing Architecture. But you didn't really need to know that!

contact

Tony Self has been involved in documentation for 30 years. In 1993, Tony founded HyperWrite, a consultancy company specializing in hypertext. Tony also lectures in technical communication and digital literacies at Swinburne University in Melbourne, Australia. He is a member of the DITA TC.

tself@hyperwrite.com
www.hyperwrite.com