DITA 101

Fundamentals of DITA for Authors and Managers

Ann Rockley, Steve Manning and Charles Cooper

Foreword by Scott Abel
Founder of The Content Wrangler

Acknowledgements

This was truly a group effort. I'd like to thank Steve Manning and Charles Cooper for all their assistance in putting this book together.

I would also like to thank Rhonda Truitt and Laurel Simmons who provided edits in a very tight time frame.

We would also like to thank the vendors for providing us with copies of their editors to use to capture examples and create content.

- Adobe FrameMaker
- JustSystems XMetaL
- PTC Arbortext Editor
- Quark Dynamic Publishing Solution for Technical Publications

For information, contact:

The Rockley Group Inc.,
Email: moreinfo@rockley.com
Website: www.rockley.com

www.dita101.com

Foreword

The pace at which technological innovation occurs is amazing. The last 20 years have been jam-packed with paradigm-shifting technological advances that have altered forever the way we create, manage, and deliver information. The personal computer, the World Wide Web, desktop publishing, touch-screen mobile phones, interactive television, social networks, and wireless connectivity have transformed not only the way consumers interact with content, but these advances have also altered the way professional communicators work.

Nowhere are these changes more evident than in the world of technical communication. Technical writers and editors have been forced – like it or not – to move to a more formal method of creating content, often for a global audience. Gone are the days of the free-for-all approach to creating technical documentation products one-at-a-time using desktop publishing tools. While this technique was the best method possible in the 80s and 90s, today, those who create user manuals, online help systems, and other types of documentation are increasingly expected to take a more formal approach to content creation, utilizing content standards like the Darwin Information Typing Architecture (DITA) - the subject of this book.

The advent of the Extensible Markup Language (XML) and rapid adoption of topic-based content standards like DITA have forced us to separate content from format and end our addiction to desktop publishing. Today, technical communicators must learn to write modular, topic-based, context-independent content using a new breed of authoring tools. It's not an easy change for many. The resources available are often poorly conceived, confusing, jargon-laden collections of information that don't make learning new skills and techniques easy. In fact, they make things much harder than they need be.

That's why DITA 101 by Ann Rockley, Steve Manning and Charles Cooper is such an important work. Simple, easy-to-understand, and loaded with practical examples that resonate with technical communicators, Rockley and team have consolidated years of experience helping folks just like you make the move from unstructured content creation to DITA. This work is the result of their efforts and a valuable contribution to the technical communication literature.

Go ahead, take a peek. It's not as scary as you might think.

Scott Abel, The Content Wrangler

Table of contents

Introduction

More and more companies are either adopting DITA or thinking about adopting it, yet surprisingly, there are very few resources written in plain English for those who want to learn about it. Many of the existing resources are very technical in nature - notably the DITA specification. Often you might find yourself asking "how do I find out more about that?" and the response is, "it is all in the spec."

And there's the problem - the DITA specification is just that - a technical specification describing the standard. It wasn't designed as a teaching tool and it shouldn't be used as one. The spec is something a DITA technologist should read, not the average writer or manager.

The good news is that we've designed DITA 101 *for writers and managers.* We've taken our years of experience helping organizations to move to DITA and distilled it into an easy-to-read and understandable format. And since the move to DITA often goes hand-in-hand with an organization's adoption of content management, we've made sure that our expertise in developing effective content, reuse, and their appropriate strategies are integrated here to give you everything you need to know to understand DITA from a author's or manager's viewpoint.

- *What is DITA* provides a definition of DITA and explains its key features.
- *The value of structure in content* explains how to analyze content for structure and uses recipes to illustrate the concepts of structure and structured writing.
- *Reuse: Today's best practice* presents the benefits of reuse and provides an understanding of all the types of reuse that DITA supports.
- *Topics and maps* takes you through the building blocks of DITA.
- *A day in the life of a DITA author* helps to illustrate how an author would develop a content outline, create topics and write structured content.
- *Planning for DITA* provides insights into how to most effectively plan for DITA and the changes in roles and responsibilities.

- *Metadata* introduces the concepts of metadata, an often overlooked and misunderstood topic, and gives insight into the metadata that DITA supports.
- *DITA and technology* introduces some of the key DITA features that authoring, content management and publishing systems should support.
- *The advanced stuff* contains our insight on such topics as domains, conrefs, selection attributes, relationship tables and specialization.
- *Appendix A and B* provide a quick reference to topic elements and prolog metadata respectively.
- *Appendix C* gives more background on XML if you would like to delve in deeper.

Examples are used throughout to illustrate the concepts.

We hope you find DITA 101 useful.

Reuse: Today's best practice

Content reuse is the practice of using existing components of content to develop new "documents." Text-based materials are the easiest to reuse. It's easier to reuse graphics, charts and media in their entirety than it is to use portions of them, but it is possible to create reusable media. We focus on text-based reuse in this book.

Most organizations already reuse content by copying and pasting content wherever they need it. This works well until the content has to be updated when it's usually very time-consuming to find and change all those places where the content has been used. Not only does this waste time but you run the real danger of missing some instances of content which can result in inconsistencies and inaccuracies.

Then again, a whole lot of organizations write content, rewrite content and then rewrite again! This may be because one of the following is true:

- Multiple people are writing the same or very similar content and do not know that someone else has already written it.
- Everyone feels that their customer and content requirements are different, so the content has to be rewritten.
- Different authors may be assigned to different media (e.g., one author is responsible for Help, while another is responsible for print and yet another for training materials).
- Some authors feel it's just too hard to find what is already written, and therefore, it's much faster to write the content from scratch.

Why reuse content?

When we write content for technical documents, particularly when we have product suites, product solutions and common functionality, there are many good reasons for reusing content.

Reusing content can provide a dramatic improvement in the way content is created in an organization. Improvements include increased quality and consistency, reduced time and costs for development and maintenance and reduced costs of translation.

Reduced development, review and maintenance

Development costs are reduced because the amount of content a author has to create is reduced. Authors do not have to research and write it again, they simply reuse it.

In addition to taking less time to create the content, less time is required to review the content. When approved content is reused, it's not necessary to review it again. This frees up reviewers to do their "real job."

When content is reused, everywhere that content is reused is automatically updated.

ROI: Is based on the percentage of reuse. Typically this is a minimum of 25%, but could be as high as 80%, depending upon how the content is reused (e.g., across types of information such as Help and print, the percentage of reuse is often very high and is also high across variations of a product).

Translation

You can significantly reduce the cost of translation through reuse. Translation memory systems (TMS) use pattern matching to match content that has already been translated so the content does not have to be translated again. Each time content is sent for translation, it's run through the translation memory tool to identify content strings (text) that have been already translated and the existing translation is reused. As you plan for reuse, you ensure that more content is reused, and therefore translation costs are reduced even further. In addition, if content has already been translated, only new or changed content needs to be sent out for translation which reduces costs even more.

Translated content can also be rapidly reconfigured and brand new information products can even be delivered from existing elements that have already been translated, without ever having to send that content to translation and pay additional costs.

The less easily measured benefits of consistent structure, consistent terminology and standardized writing guidelines that reuse requires also help to reduce the cost of translation.

Often, a large cost in translation is in reformatting content. Frequently, content must be converted from the original source format (e.g., Help, FrameMaker, HTML) to RTF (rich text format) before it can be translated. When it's converted, it loses much of its formatting. When content is in XML (e.g., DITA), it's easy to automatically reformat content, regardless of language.

ROI: One of the biggest areas of ROI for DITA and reuse is with translation. If four or more languages are translated, typically all costs can be recouped in less than 18 months (including the cost of purchasing a content management system). Some specific areas of translation ROI include:

- The cost of translation is reduced by the percentage of reuse (typically a minimum of 25%).
- The cost of reviewing the translated content is also reduced by the percentage of reuse.
- Desktop publishing/post translation formatting is typically reduced by 30-50%.

Increased consistency

When there is no reuse, the chances of inconsistencies in content increases, either because the content has been rewritten by many people, or because it has been copied and pasted and some of the occurrences of the content have not been updated properly. Often, when content is copied and pasted, the versions of the content begin to diverge over time.

When samples of materials are examined, examples where content is similar, but not exactly the same are found. On average, five to six variations of content are found. The worst we have seen is 56! Yet, nine times out of ten, when we and our client really look at the information, we realize that the content could be identical.

When content is written once and reused many times, it ensures that the content is consistent wherever it is used. This consistency leads to higher quality content.

ROI: ROI is based on cost avoidance.

Rapid reconfiguration

Reusable content is modular content (small self-contained topics that can be used in combination with other topics). In today's rapidly changing world, products and customer requirements are constantly changing. Modular reusable content makes it easy for organizations to rapidly reconfigure their content to meet changing needs. You can easily change the order of modules, include new modules, exclude existing modules and use modules to build entirely new information products to meet new needs.

ROI: The ROI for rapid configuration is the opportunity cost.

Real world examples of the benefits of reusing content

High Technology

A high tech firm found an error and documented how to avoid the problem. The change was documented in a technical bulletin, but it never made it back to the site alerts or regular documentation. All of a sudden, customer systems started failing, resulting in millions of dollars of lost revenue and very unhappy customers! The problem was most customers had not read the technical bulletin and because the information did not show up anywhere else, they were not aware of it.

Medical Devices

A medical devices firm was hit with a $10,000,000 lawsuit. Content in the Physician's Guide was different from content in the Patient's Guide which was different from what was on the Web site. A doctor provided guidance to a patient who went back to the documentation to remember what to do. The documentation was wrong and the patient ended up being hospitalized. The medical devices firm lost the lawsuit.

Processes of reuse

There are multiple methods of reusing content.

Opportunistic reuse

Opportunistic reuse occurs when the author makes a conscious decision to find an element, retrieve it and reuse it. Opportunistic reuse is the most common form of reuse. However, opportunistic reuse relies on the motivation of the author to know that reusable content exists and to go and find it.

Any content can be used in an opportunistic reuse situation. In some ways, opportunistic reuse is a replacement for "copy and paste." However, opportunistic reuse is not copy and paste because it's actually a "pointer" to the source content.

Systematic reuse

Systematic reuse is planned reuse. If you have a component content management system, it can automate this type of reuse. Specific content is identified as reusable in a particular place and is automatically inserted. This can be done at authoring (e.g., as a specific template is used and content identified) or at delivery (e.g., personalization).

Systematic reuse can significantly increase the percentage of reuse because reuse can be planned and ensured. However, systematic reuse requires a greater level of information architecture to identify and implement the opportunities for reuse.

Types of reuse

DITA was specifically designed for reuse. Topics, sections, paragraphs, sentences, or even words can be reused. The types of reuse will be described here and how you accomplish reuse with DITA will be described in more detail later in the book.

DITA supports a number of different types of reuse including:

Topic-based reuse

Topic-based reuse is the process in which authors create content as individual topics. Topic-based reuse is at the heart of DITA. "A topic is a discrete piece of content that is about a specific subject, has an identifiable purpose, and can stand alone."[1] Authors assemble topics to create information products and topics are assembled together into information products using DITA maps (see "Map" on page 41). For example, the third recipe we showed you for pumpkin cheesecake could be assembled as a map of three topics, each of which is a recipe.

Figure 1: Topic-based reuse within a recipe
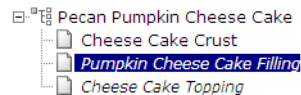
Topic-based writing can sometimes be the hardest thing authors need to learn to do when moving to DITA. Guidelines for how to create topic-based content can be found in "What is a topic " on page 46.

Fragment-based reuse

Pieces of a topic, like a paragraph or a section can also be reused. This is referred to as fragment-based reuse because a fragment (piece) of a topic is used. A fragment may include a title like a topic does, but it also may be a paragraph or even a sentence. In the past, to reuse a small piece of information like a paragraph, the content had to be chunked out and saved as an individual piece/file of information. DITA makes it possible to point to any structure in a topic

---

1.   http://en.wikipedia.org/wiki/Topic-based_authoring

(e.g., p [paragraph], stepxmp [step example]) and identify that we want to reuse that information. This makes small fragments of information much more manageable.

Fragment-based reuse can either be opportunistic (known content exists so the author takes the opportunity to find it or reuse it) or systematic (planned reuse). It's not possible to have writing best practices for opportunistic reuse, but it is for systematic reuse. While any element in any topic can be pointed to and identified for reuse, it makes a lot of sense to group together content planned for reuse (systematic reuse). For example, if a author wants to reuse a warning, the author could point to it in the topic that includes it, but the next author who wants to reuse that warning would have a very difficult time knowing where that warning is located to reuse it. It makes more sense to group together commonly reused content fragments into a single topic for ease of search and retrieval. Or in the case of the recipes, you could group together the descriptions of different ingredients such as couscous which is a type of pasta.

Fragment-based reuse is supported in DITA with conrefs (see "Conrefs" on page 82). For example, we could create a Pasta topic that provides more detail on different types of pasta and reuse the fragment on couscous into the recipe.

·<enter·topic·title·here>Pasta¶

▪¶
Couscous·is·a·very·tiny·pasta·that·originated·in·North·Africa.¶

¶
Penne·are·tube-shaped·pasta·with·angled·ends·cut·to·resemble·a·quill·or·pen·point.·Penne·pasta·means·"pen"·in·Italian.·¶

¶
Farfalle·are·rectangular·or·oval·pieces·of·pasta·that·are·pinched·in·the·middle.·They·look·like·bow·ties.·Farfalle,·means·"butterfly"·in·Italian.·¶

¶

\<enter task title here>Mediterranean Couscous Salad

Sunny flavors of tomatoes, olives and oil team up with couscous (available in most supermarkets) to create this summer salad.

Pre-requisites

1 ¾ cups	Water	435 ml.
½ teaspoon	Salt	2 ml.
1 ¼ cups	Couscous	300 ml
1/3 cup	Chopped green onions	75 ml
2	Tomatoes, diced	2
1 cup	Diced feta cheese	250 ml.
½ cup	Black olives, pitted & quartered	125 ml.
2 tablespoons	Each chopped fresh mint and oregano	25 ml
1/3 cup	Good quality olive oil	75 ml
¼ cup	Lemon juice	50 ml
	Pepper to taste	

Context Information

Couscous is a very tiny pasta that originated in North Africa.

1. In a saucepan, bring water and salt to boil; add couscous and stir.

2. Cover and remove from heat; let stand 4 minutes. With fork, stir to separate grains.

3. In large bowl, combine couscous, green onions, tomatoes, feta cheese, olives, mint and oregano.

4. Whisk together olive oil and lemon juice; pour over salad and toss to combine.

Figure 2: Couscous fragment reused in the Mediterranean Couscous Salad recipe. Note that the authors structural cues like "task title" and "context" have been deliberately left in the image for understanding.

Filtered reuse

In filtered reuse, authors provide variants for a specific chunk of information in a single component. The variations are identified by conditional tags or attributes. When the topic is "published," the different variations are published as required.

Filtered reuse is very valuable in multichannel publishing, audience variants, product variants, region variants and so on. It's an intuitive way for authors to keep all variations together for ease of writing and review.

Also, all the variants must be written and structured in such a way that they are consistent with other content they will be used with. For example, if a variant element includes content for training, a user guide and for online Help, the content that is published to Help must be structured in such a way that it is consistent with the other content written for the Help file. The same goes for training and for the user guide.

We sometimes refer to the creation of filtered reuse as the "building block" approach. That's because there's a core set of content that's applicable in all uses and variant content that *builds* on the core content that is only applicable in certain situations. To create content for filtered reuse:

- Identify the core content (the content that is applicable for all uses).
- Identify what has to be added, what has to differ to meet other needs.
- Make sure that when the content is filtered (e.g., the irrelevant content is filtered out), the content still flows.
- Tag the elements, indicating where they belong.

Use of the building block approach allows authors to create all the content for a reusable element at the same time; in this way, all reusable content for an element resides together and can be reused in its entirety.

Filtered reuse is often thought of as conditional reuse, (e.g., like conditionals in FrameMaker) and is accomplished using Selection Attributes in DITA (see "Selection attributes (conditional content)" on page 84). Attributes is another name for metadata. (Refer to "Metadata" on page 61 for more information on metadata.) In FrameMaker you would create a condition tag, in DITA you create a tag (attribute) for selecting the content to hide or show when content is published.

For example, we could decide to filter the description for couscous and only display it for novice cooks. In this case, we have created an attribute (tag) called Audience that has a value of Novice that lets us show the description for couscous in the Novice version of the recipes, but hide the couscous description for more experienced cooks.
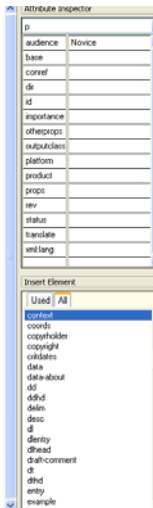
Figure 3: Illustrates the Audience value of Novice applied to the couscous description.

Variable reuse

Variable reuse occurs when a variable is set up that can have a different value in different situations. For example, the name of a product might be one thing in North America and another in the European Union. Variable reuse is very useful for small pieces of information.

Variable reuse becomes valuable when there are only slight variations in content (such as product names in different regions), but otherwise the rest of the content is identical.

When creating the values for variables, remember that the value has to transparently fit into the content so that the content flows like a regular sentence that does not include a variable.

Looking closely at the recipes, it's obvious they are written by different people, particularly when looking at the measurements. There is tablespoon, tbl and tbsps. These should be consistent. You could use a variable to control the terminology used for measures. Variables are also supported with conrefs. Refer to "Conrefs" on page 82 to see how variables can be accomplished using DITA.