
# Preparing for DITA: What you need to know

**White Paper**

August 16, 2005

Prepared by: The Rockley Group Inc.

Phone: 905-415-1885

Fax: 905-415-3413

Email: moreinfo@rockley.com

Website: www.rockley.com

For information, contact:

The Rockley Group Inc.,
445 Apple Creek Blvd. Suite 120, Markham, ON L3R 9X7

Email: moreinfo@rockley.com
Website: www.rockley.com

# Table of contents

## Introduction

If you're involved in information development, chances are the Darwin Information Typing Architecture (DITA) is a part of your future. DITA is an exciting advancement that has the potential to help Extensible Markup Language (XML) become the standard approach to authoring technical content.

This paper (the first in a two part series) describes the features of DITA that you need to know about, in order to determine if DITA is the right solution for you. Please note, that this paper is not an a technical "how to." Nor is it an exhaustive description of all features in DITA. Instead, we will attempt to rise above some of the technical jargon that accompanies most descriptions of DITA and explain why DITA merits the hype.

For those of you more interested in the "how to" of DITA, the second paper in this series will delve into more of the details of analysing and adapting DITA to meet your information development needs. That paper, co-authored by The Rockley Group and BlastRadius Corporation, will be available in late September.

## A very brief history

DITA, or Darwin Information Typing Architecture, was developed primarily by IBM in a response to the changing needs of their business. Those needs are the same needs that we all face:

- Figuring out how to get products to market faster
- Finding ways to reduce unnecessary expenses
- Delivering content in an increasing number of output formats
- Finding ways to react faster to changing demands (more flexibility)
- Increasing the effectiveness of content

Changes in corporate goals, changes in technology, changes in customer expectations and needs all have to be met. DITA is the mechanism with which IBM chose to meet those needs.

### OASIS

Recognizing that DITA would benefit writing departments everywhere, IBM has passed the standard to the Organization for the Advancement of Structured Information Standards (OASIS) a not-for-profit, international consortium that drives the development, convergence, and adoption of e-business standards. As an OASIS standard, DITA is at version 1.0, with 1.1 in development.

### Design goals

Before discussing the details of DITA, it is useful to understand some of the goals that lead to the development of the standard.

*Move away from focus on books to multiple formats and outputs*

One of the largest impacts of technology on information development is the addition of so many new formats for delivering information. No one just delivers a user guide (book) any more. There is increasing need for information to be delivered in multiple formatss. While some improvements have been made, many of the book-based technologies in use today are not efficient tools for creating multiple outputs from a single source of content. Format conversions have been simplified, but working in a book-based paradigm has made it difficult to support other information types that do not need the same kinds of book structures. The standards that have been available to date, like DocBook, have been book-based.

*Move away from SGML to XML*

IBM had been using SGML for some time but recognized that XML, with its' stated focus on Internet applicability, was a better option. Web-based formats and delivery have become crucial tools for delivering effective information to users.

*Move towards the trend to minimalism*

The goal of information development and delivery should be "the right information, at the right time, in the right format, to the right person." For IBM, that meant reducing information "glut", lessening the volume of irrelevant information presented to users, and focusing on providing only the information that users *need*. This approach reduces both the time it takes to create and maintain information, facilitates quicker information delivery, and reduces the effort required to keep information up-to-date.

*Provide more flexibility in structures and move away from "monolithic" DTDs*

The trend in the past has been to create a Document Type Definition (DTD) that focuses on the needs of single departments and specific information products they produce. This approach, however, can have a negative impact on wide-scale reuse, as content created by one department may be difficult to reuse in other departments. DITA is intended to provide a mechanism that establishes a clear base of common structures, while making it easier to create specific structures needed by different departments.

*Support maximum REUSE*

Reuse is today's best practice for information developers. DITA was developed specifically to promote content reuse and reduce redundant information.

## Key concepts

A starting point for understanding DITA is to examine the key concepts of the standard and the advantages they offer.

- XML
- Modular content objects
- Topics
- Reuse
- Specialization
- Topic maps
- Conditional text

### XML

XML has become the technology of choice for complex content management. DITA, is a specific implementation of an XML markup language. The advantages of authoring with XML over traditional word processing and report writing tools are well documented, but are worth repeating.

#### *Extensible*

In XML, you decide the tag names that you'll use in your documents. Using XML, you can create semantic tags (tags that have meaning), rather than generic tags. You apply tags that describe the content of the information, not the formatting. For example, typically in Microsoft Word there is a generic tag called "Normal" that you apply to information that you want to be formatted in a certain way. In XML, you can create semantic tags to identify individual components of information (e.g. "introduction", "title", "product description", "process", or "objective". Semantic tags provide human- and machine-readable metadata (information about the content they enclose), and can be automatically interpreted (rendered) for display in many ways. DITA uses a combination of semantic elements to describe specific structures as well as more generic tags (borrowed from HTML) where semantic names may not be necessary.

To increase the preciseness of metadata, XML allows authors to define attributes that describe in specific detail the elements of content contained in a document. Similar in syntax to HTML attributes ("color='red'"), XML attributes are defined by you, to provide whatever additional information is required to identify the use of the information. Attributes can be used, among other things, to identify information that is specific to format, product, user, or use. DITA makes good use of attributes for both content and metadata.

#### *Hierarchical*

The basic physical structure of traditional word-processing and desktop publishing tools is the paragraph. That's the block-level construct that you apply styles to. While such an approach does make it possible to format content manually, it is not an effective approach for identifying and manipulating the structure of content. There's no easy way to identify the relationship between the paragraphs to form larger structures, like sections and subsections

DITA, as an XML implementation is hierarchical in nature, with structures of content having specific boundaries – start and end tags – as well as defined substructures. This hierarchical approach helps give structural context to elements of content and makes XML a much easier mechanism for manipulating content. (In XML, when you select an element, all of its child elements are automatically selected. In word processors, you must manually select all of the paragraphs that comprise a chunk of content.)

*Structural*

XML requires a DTD (Document Type Definition) to support the development and management of content. The DTD is like a structural template: it explicitly defines the structure of the content. This explicit structure ensures that authors can only enter content that follows the structural rules. That is, all of the required pieces of information are in place, are used with the correct frequency, and appear in the correct order. There is no possibility of entering invalid elements in the structure. This will assist authors in writing rapidly and eliminates many structural and consistency errors. DITA includes a complete DTD that defines a complete collection of elements and attributes.

*Separation of content and format*

Authors using products like Microsoft Word are accustomed to applying style tags that define the "look-and-feel" of the content they are creating. In an XML authoring environment, authors apply tags to define the meaning or purpose of the content while style sheets associated with the content automatically define the look-and-feel of the output. For example, stylesheets can be created to control the way content appears on paper, while another stylesheet can be created to control the way the content appears in a web browser, and still other stylesheets can be created to support the desired layout and formatting of the same content appearing in a slide show, or an annual report. The end result is that multiple stylesheets can be applied in sequence to the same content to create multiple formats automatically or as needed.

The look-and-feel is defined by the appropriate stylesheet selected in the final production process, or at any time in the authoring/review cycle. The DITA toolkit includes a variety of stylesheets for different outputs. You can use these as a starting point to create your own stylesheets.

## Modular content objects (topics)

A key feature of DITA is that information is organized (and) stored as modular chunks of content. The chunks, or topics as they are known, can be reused as building blocks of content. Topics can also be "typed". That is, you can create different types of topics with a predefined structure that is appropriate only to that topic type. For example, a concept has a different substructure than a task.

All topics in DITA are built on a single model of a generic topic. This generic type defines the elements that are common to topics of all types. However, DITA recognizes that different topic types need different substructures (for example, a procedure requires a different substructure than a concept) and allows for these variances.

As a starting point, DITA version 1.0 includes the following specialized types:

- Concept
- Task
- Reference

**Specialization**

Not all information needs can be satisfied with DITA (generic) Topic, Concept, Task, or Reference types. DITA developers recognized this and included functionality to help users create their own, custom topic types. The process of customization is called specialization.

DITA DTDs and stylesheets are designed to provide users with a mechanism for creating specialized topics. This functionality allows you to add elements to existing topics that supplement base DITA structures. This functionality comes in handy when you need to identify elements of content that DITA base models have not included. You can also add new topic types as needed.

Specialized types define additional elements that are specific to the type of content they will be used to mark up.

Support for specialization is a key differentiator when comparing DITA to other standards. Typically, standards have not directly supported customization.

**Reuse**

Support for content reuse is a key goal for DITA. Topics written as self-contained (modular) chunks can be reused in one or more places in a single information product or in one or more places in multiple information products. Topics can be combined in different information products and in different combinations to provide needed variations.

DITA also provides a mechanism called "content references" (con-refs) that allows users to identify content within topics for reuse. This allows a piece of content to physically reside in one topic, where it will exist and be maintained in context, but still be reused in others.

**Topic maps**

Having a collection of topics for your information product isn't very effective without some sort of structure to organize the topics in a logical, hierarchical, and accessible way. We rarely distribute individual chunks of information. Instead, the chunks are combined to create output products, like online help systems, product user guides, FAQs etc.

In DITA, Topic Maps define the organization, order, and hierarchy of topics in an information product. (They resemble a table of contents in an online help system.)

Topic Maps provide pointers to topics: the topics are not embedded. This allows the same topic to appear in different places in a single information product or series of information products. Topics can also appear in different Topic Maps.

The map structure also allows you to add information about topics in the map to provide additional context processing the topic to output.

**Conditional text**

Conditional text is another method of adding reuse to content, but within topics. It's been a popular feature of several text-authoring tools for some time. Using DITA, you can add condition indicators to elements of content to identify the context in which that content should or should not appear. For example, elements could be identified as being appropriate for reuse in online help, but not in print, in the French version, but not the Chinese, or in two specific products in a suite of ten products.

## DITA as a gateway to XML

Statistics have shown that XML adoption has been greatest in industries where there was a defined standard in place. An existing standard can provide short cuts to adopting a technology like XML. DITA has the potential to provide a quick opportunity for many companies to take immediate advantage of the many benefits provided by both XML and content reuse.

By adopting a standard you can reduce the effort required to code and test a DTD or Schema. For companies with limited IT resources or those lacking in an experienced talent pool, adopting standards can yield tremendous savings in time, resources, and money. DTD creation is a skill of its own, with a defined learning curve. By adopting an existing DTD you can take a shortcut and benefit from the DTD building expertise of those who developed the standard.

The same is true for developing stylesheets to convert XML content into other formats, like HTML and PDF. The XML stylesheet language, XSL, can be used to create some pretty complex output. However, the more complex the output needed, the more complex – and expensive -- the stylesheet you create will be. XSL is powerful and is complex. Complex stylesheets require considerable XSL and programming knowledge and experience.

Alternatively, many standards come with bundled stylesheets for many different output types. The creators of DITA have recognized that a DTD alone only addresses the development side of the content lifecycle and have provided stylesheets as part of the DITA toolkit to help users quickly develop user-ready outputs from their DITA content.

## DITA and tools

As the current "hot" technology, tools vendors are quickly adding support for DITA to their own offerings. However, in reviewing tools for use with DITA you must be careful in assessing the real support they might offer. As an XML-based technology, any XML tool can be said to support DITA. That is, if the tool is capable of working with standard XML structures and standards, it will work with DITA. But as a differentiator, you need to determine if the tools up for consideration offer any useful, additional functionality. For example, an XML editing tool might provide menu functions to help you build a Topic Map, perhaps with an "Explorer"-type mechanism and drag-and-drop interface, thus removing the need for you to manually build the Topic Map by typing in file names. Other software tools might offer functionality that allows you to quickly and easily insert different topic types without having to select them from a list of elements.

It is therefore very important to ask software vendors "How do you support DITA?"

## Is DITA right for you?

That's the key question, right? Will implementing DITA in your department or organization be the path to meeting your goals for lower costs, faster turnaround, or whatever? It can be a difficult question to answer; it really depends on your content and what you need to do with it.

For departments creating online help or help-like information, DITA certainly looks like a good fit. The base types of Concept, Task, and Reference map very easily to the standard topics types that are default in many online help systems. It's easy to make a mental leap from implementing online help in a help tool to implementing online help in DITA.

If your focus is book-based information, it's not as easy to see DITA as a solution. DITA doesn't currently incorporate structures like front matter and introductions that we expect and need in books. If your materials are not book-based, you may find that your content doesn't fit the DITA structure. Marketing materials, annual reports, pharmaceutical labels, product data sheets and many other types of content may require a significant investment in time, effort and resources to specialize DITA. In these cases, a customized DTD might be a better solution. So, the question is not "Is DITA right for me?," but rather, "How much customization will I have to do to make DITA right for me?" By design, DITA is intended to be extensible to meet any content need! You must decide if the returns will justify the customization effort or if it is more effective to create a custom DTD.

To really figure out if DITA is a good solution for you, and to figure out how much customization will be required, you must begin by having a thorough understanding of your own content, but not from the usual perspective of content as information. You need to develop a structural understanding of your content. What are the structures in your information product? If you create user guides, what is the structure of a procedure? Do you have one type of procedure or several? What other structures do you need? If you create brochures or product descriptions, what are the common elements of content? How are they structured?

If you create a model of your content, you can compare it to the structural definition of the topics in base DITA. That will help you to identify where DITA fits exactly and where you must create specialized elements or topics to add the structures you need.

## What is content modeling?

One of the most critical activities in developing a content management strategy is understanding the content that is created and delivered to customers. The understanding must be "360". That is, you must understand the full lifecycle of content: how the content is created, how the content is delivered, how it is used, and when it is no longer useful.

During content modeling, you determine the elements required for each information product (or output) and how each information product will be designed for optimum usability and reuse. You decide which information products you will produce and the information they will each contain. You must identify the metadata and content standards to support them. Thus, the content models become the road map for your project.

In a content model, you break information down to the element level (e.g., section, paragraph, sentence). You identify how the elements will be stored, how they will be shared, how they will travel through workflow (for authoring and review) and how they will be compiled into new output products.

The power of content reuse lies in effectively reusing content elements—whether they're paragraphs, procedures, or sentences—over and over again. Information models identify all the required elements and illustrate how to structure and reuse them. Developing a model of your content puts you in a good position to evaluate the effectiveness of DITA and the amount customization that would be required for your organization. It also provides a focal point for comparing DITA against other standards that are available.

## Alternatives to DITA

You may need to evaluate DITA against other standards. The buzz about DITA in the information development community has been so great that many writing departments with custom DTDs or those using other standards have been asked by their management to evaluate (or on their own initiative are evaluating) DITA to determine if there is a case to me made for switching. Or, your information models may be so different that you feel the cost of customizing DITA to support them will be prohibitive.

There are many XML applications and standards available, too many to list here. A very good list can be found at http://xml.coverpages.org/xmlApplications.html.

## Summary

DITA is an exiting technology for information development. It incorporates many of the current best practices:

- XML-based
- Modular content
- Reuse

DITA isn't right for every purpose, but the only way you'll know if DITA is right for you is by examining your content critically and analytically, creating structural models of your content, and understanding the how your content fits into your organizations content lifecycle.

For more on DITA, the second paper in this series will delve into more of the details of analysing and adapting DITA to meet your information development needs. That paper, co-authored by The Rockley Group and BlastRadius Corporation, will be available in late September.

## Appendix A: Company profile

The Rockley Group helps content managers and authors meet the increasing demands of creating, distributing and managing the content they create. Our team of experienced analysts (link to your team page) bring a wide variety of expertise to the table and can help you avoid expensive pitfalls. Organizations of all sizes -- from small, privately-owned firms to multi-national Fortune 500 companies -- trust us with their most important content projects. We serve clients in the Financial, Life Sciences, and High Technology industries, as well as others in the Communication, Marketing and Retail Sales markets. We've developed content reuse solutions that reduce the cost and effort to produce complex information products including: marketing collateral, software documentation, online help, customer support materials, human resources content, as well as regulatory documents for pharmaceutical and medical device manufacturers.

The Rockley Group was established in 1995 to serve the information creation community. Founding president Ann Rockley has more than 20 years' experience in online documentation, web design, instructional design, enterprise content management and content reuse (single sourcing). But The Rockley Group is more than a one person shop. Our team of experienced analysts, instructional and online content designers, information architects, project managers, information technologists, editors, writers, programmer analysts, and engineers provide our clients with the skills necessary to deliver content management solutions that work.

The Rockley Group has been in the forefront of content development research, and has assisted in the evolution of industry standards for online documentation, web-based instructional design, content reuse, and enterprise content management. We're passionately committed to discovering innovations in the field of content design and management. This commitment is evident in our belief in education as part of their responsibility to the information creation community. Senior members of The Rockley Group team regularly teach university courses and seminars, speak at industry conferences, publishes articles, and present workshops and papers around the world related to XML, enterprise content management, e-learning and single sourcing. And, several members of our team serve on committees designed to establish international standards for content management, information design, and documentation development.

Our commitment to education doesn't stop there. We've documented our methodologies in our recently published book, *Managing Enterprise Content: A Unified Content Strategy* (ISBN 0735713065) New Riders Publishing. Authored by Ann Rockley with senior team members Pamela Kostur and Steve Manning, the strategies introduced in *Managing Enterprise Content* have been heralded as content management best practices and have been adopted by some of the world's largest organizations.

## Appendix B: Customer list

### Financial

- Bank of Canada
- Bank of Montreal
- Bank of Nova Scotia
- Canadian Imperial Bank of Commerce
- CGU Group Canada Ltd.
- Citibank Canada
- Coopers & Lybrand
- Deloitte & Touche
- The Investment Funds Institute of Canada
- The Dominion of Canada General Insurance Company
- Manulife Financial
- Norwest Services
- Ontario Municipal Employees Retirement System (OMERS)
- OPSEU Pension Trust
- Sun Life of Canada

### Life sciences

- Agilent
- Anthem Blue Cross/Blue Shield
- Guidant Corporation
- IDX Systems Corporation
- Eli Lilly
- International Aids Vaccine Initiative
- ISG Technologies
- Johnson & Johnson
- Medtronic Inc.
- MDS Sciex
- Nellcore, Puritan, Bennett, Mallencrodt
- Visible Genetics

**High tech**

- Aliant Inc.
- Bell Sygma
- Brain North America Inc.
- Cisco Systems
- Compaq
- Delano
- Digital Cement
- Hewlett-Packard
- Hummingbird Communications
- Intel
- Lexmark
- Nortel
- Ontario Systems Corporation
- Promis Systems Corporation
- Rogers Communications Inc.
- Sasktel Mobility
- Symantec
- Texas Instruments
- Watchfire

**Other**

- Canadian Standards Association
- Citizenship Immigration Canada
- Dofasco Inc.
- Dynegy
- Environment Canada
- GO Transit
- Inco Ltd.
- Pacific Northwest National Lab
- Purolator Courier
- Sears Canada Inc.
- Schlumberger