



The Role of Content Standards in Content Management

White Paper

February 15, 2005

Prepared by: The Rockley Group Inc.

Phone: 905-415-1885

Fax: 905-415-3413

Email: moreinfo@rockley.com

www.rockley.com

©2005 The Rockley Group Inc. All rights reserved. The whitepaper can be posted in its entirety on a web site for educational or knowledge transfer purposes, commercial or otherwise. No other use of this whitepaper except for brief quotations in articles or presentations is authorized without written permission.

For information, contact:

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

Email: moreinfo@rockley.com

Website: <http://www.rockley.com/>

Table of contents

Introduction	1
Why go to standards? Why not?	2
Regulatory compliance	2
Industry conformance	2
Information interchange	2
Technological opportunity	2
Where do the standards fit in the process?	4
Effect on delivery	4
Effect on authoring	5
Standards and content models	7
What is content modeling?	7
Content models and standards	8
Popular standards	9
DocBook	9
DITA	10
SCORM	10
SPL	11
Others	11
Summary	12
Appendix A: Company profile	13
Appendix B: Customer list	14

Introduction

With the increase in popularity of XML has come an increase in the availability and popularity of document standards. Technical publications departments are busy learning all they can about DITA (the Darwin Information Typing Architecture), or Docbook, or ATA, or S1000D. In pharmaceutical industries, they're concerned about SPL (Structured Product Labeling). Learning groups are dealing with SCORM (Sharable Content Object Reference Model). There's lots to learn.

Note that in this context standards are defined content structures that you can adopt for some part of the content lifecycle. Many of these standards are implemented using XML.

Standards are being implemented for many reasons. This whitepaper explains the value of standards and, more importantly, their effect on your content management strategy.

(For more information on standards and the impact on content processes, see the March issue of *The Rockley Report* (www.rockleyreport.com).

Why go to standards? Why not?

We should all be used to standards to follow. We've all had style guides to help constrain our structural or language choices. But for the most part, they've been internal standards, created within a company or department to be used internally. Now, we are frequently faced with having to adopt external standards—created by people in other organizations—to do our jobs and to compete in the global marketplace. There are many good reasons to adopt content standards, ranging from necessity to opportunity. Some of the reasons include:

- regulatory compliance
- contractual requirements
- interchange
- technological opportunity

Regulatory compliance

Some industries must comply with regulations set up by agencies, usually governmental, to control content. (For example, pharmaceutical and medical device companies doing business in the United States must comply with the regulations of the Food and Drug Administration.) Failure to do so can result in lost opportunities and potentially legal action.

Industry conformance

Some large industries (like the auto and aerospace industries) have set up content structures as a means to standardization. For a manufacturer adopting the standard, content standards become a mandatory format for anyone delivering information to them. It provides consistency and more importantly predictability. Companies can predict the structure of the information being provided to them. This opens up opportunities for automatic processing of the content. The first industry content standard (CALs) was created by the US military to bring consistency and predictability to the thousands of vendors who supply information along with their products.

Information interchange

Some standards have been developed as a means to share or exchange information among companies or groups with common interests.

Technological opportunity

Some standards have been developed and/or adopted as an opportunity to take advantage of technology or new processes (like reuse). Docbook has provided a quick opportunity for many companies to take immediate advantage of the capabilities of XML. DITA has helped companies to take advantage of both XML and reuse. 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 technology, like XML. 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 starting knowl-

edge, this can be a tremendous time saving. DTD creation is a skill of its own, with a definite learning curve. By adopting an existing DTD you can take advantage of someone else's DTD building expertise.

The same is true for stylesheets to convert from XML to other formats, like HTML and PDF. The XML stylesheet language can be used to create some pretty complex output. However, the more complex the output, the more complex the stylesheet required to create it. The XML stylesheet language, XSL is powerful and is complex. To create complex stylesheets really requires considerable XSL and programming knowledge.

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

Where do the standards fit in the process?

In order to understand the impact a standard might have on your organization and processes, it is important to understand where the standard fits in the lifecycle of content. There are three possibilities: in the delivery phase, in the authoring phase, or in both.

- **Delivery phase**

When you choose to “adopt” a standard, it does not necessarily mean that you must author in that standard. Some standards can be treated like HTML. That is, you can author in a format of your own choosing—one that makes production sense for your environment—and convert/transform the content into the appropriate output form. In this case the impact of the standard is really in the delivery phase. SCORM is an example of a standard that does not define an authoring template. It does not define the structure of learning content. Instead, it defines the structure and data of a “wrapper” that can be read by SCORM-compliant learning tools. (See SCORM.)

- **Authoring phase**

Some standards impact your processes in the authoring phase. DITA, for example, is an authoring standard. It defines content structures that authors use when creating or modifying content. Authors use the DITA DTD to create information objects that are then combined and output into another format, like HTML or PDF.

- **Both phases**

Some standards can be used for authoring and also delivered to users. For example, choosing to adopt Structured Product Labels for drug labels does not mean that a pharmaceutical company must use the SPL DTDs for authoring. They can author in a format of their own choosing—one that makes production sense for their environment—and convert/transform the content into SPL for delivery. However, they can also author using the SPL DTDs and deliver SPL files.

The phase in which the standards are applied to your process will determine the overall impact on your processes.

Deciding how to implement depends on why you are moving to the standard. For example, adopting a standard for information interchange means that you will probably deliver documents marked up with that standard.

Effect on delivery

The effect of selecting specific standards doesn't really have a major effect on your process. It represents another format the you must convert or transform to. For specialized standards, like SCORM, there are usually tools available that can create the SCORM output using an “export” like feature of the tool. For other output standards, you will need to create output scripts to transform your content into the output type.

The biggest factor in ensuring that you can output to a specific standard is to make sure that your content includes all of the content elements required for the standard and that you have defined your content in such a way as they can easily be picked out for transformation. That is, if the output standard specifies that date element exists in a certain place in the output structure, you have to have the appropriate date in your source content and it must be identified in such a way (with styles or tags) that your conversion scripts can pick up the date and output it in the correct place.

Effect on authoring

Implementing a standard in the authoring phase has a greater impact on your processes and your people. Using the structures (DTD/Schema) defined by the standard can have advantages and disadvantages.

The obvious advantage of authoring in the standard's structure is that you have a much greater chance of including all of the pieces of content prescribed by the standard.

However, using the standard's DTDs for authoring may reduce the effectiveness of the authoring process. You will need to plan for training to assist your authors in effectively authoring with the new standard.

Optional vs mandatory

Standards tend to be written to be as broad a solution as possible. That is, there are usually a core set of mandatory elements that must be included; then there are a large number of optional elements, added to the standard to try to give it the widest possible applicability. The end result is a standard that is not usable "out of the box". You need to put considerable effort into determining which of the optional elements pertain to you. Then you need to put an equal amount of effort into "hiding" the optional elements you don't need from your authors.

Some standards are written with the expectation that you will provide some customization to reduce their complexity or to extend their capabilities to meet your needs. For example, DocBook has hundreds of defined elements¹, but it also comes with comprehensive instructions on how to create a customized layer, to focus the DTD on only those elements that you need to support your structures.

You should be prepared to implement standards with customization layers or other controls to ensure the authoring process is as efficient as possible.

Whose lexicon?

When you implement a standard in authoring it usually means that you must also adopt someone else's tag set. That means, your authors must learn and use someone else's terminology. This adds to the learning curve in adopting the standard.

One of the great advantages of XML is that the tag set is user-defined. When implementing your own structures, you can choose the names of the elements (tags) to suit the language/terminology of your customers and authors.

1. There is a simplified version of DocBook with subset of the elements in the full DocBook DTD.

Granularity and reuse

If XML is the fastest growing technology in information development, then reuse is the fastest growing technique. Information delivery has changed so rapidly, providing so many new options, that information is rarely applicable to a single information product. Even single documents (like a user guide) are frequently delivered in multiple output formats. More likely, you are delivering different versions of documents with a mix of common and unique content across them all, or different types of information products that share a core set of information. A “write one, use many” implementation is the most effective process for reuse.

This can impact your implementation of a standards, as it was unlikely that an industry standard was written with your reuse needs in mind. (The exception here is the DITA standard, which was written specifically to support modular reuse.)

Granularity and collaborative authoring

The “one writer, one document” style of authoring is becoming less common in information development. The pace of updates, coupled with the reuse strategies means that, more and more, the books are created by a collaborative team of authors. And just as the granularity of a standard may not fully support your plans for reuse, it is also true that a standard may not support your plans for collaborative authoring. The granularity supported by the standard may not support the specific chunks of content that you need to create, distribute, and track.

Standards and content models

One of the least understood activities in content management is the relationship between implementing standards and content modeling. Many people think that by implementing a standard, which embodies a specific content model, they do not need to do their own content modeling, they can just adopt the content model in the standard. This is not really a good idea. In fact, the effective implementation of any standard really requires that you develop a content model before choosing a standard.

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 life-cycle of content: how the content is created, how the content is delivered, how it is used, and when it is no longer useful. It is your content.

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 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.

The process of content modeling involves identifying all the information requirements for a particular project (or for an entire organization) up front, deciding which information belongs where, then building a model that illustrates how content elements will be compiled to form each information product. Authors refer to models to determine which elements are required for each information product and to determine which elements are reused.

The content modeling process forces you to consider all information requirements and to assess what information is available to fulfill those requirements. You need to consider who produces the content, who uses the content, and how you can use the content in more than one place.

The content model becomes the “catalog” of all information products produced within an organization, and outlines the necessary content elements for each of them.

Content models and standards

A standard content model may not fit your content, nor may it support your needs for reuse and collaboration. That is why content modeling—and specifically creating formal content models—is important as a first step to implementing a standard.

First, the content modeling exercise will give you a complete understanding of your content, from both the creation and delivery perspectives, and put you in the position to be able to fully assess the impact of implementing the standard on your environment.

Secondly, the content model will help you to fully understand and define your opportunities for reuse. You will understand the information that you create and author, the elements that you store and track, and how those elements map to the specific information products that you create.

Third, you will be able to truly assess whether the standard meets your full set of requirements. Modeling will enable you to determine how you need to extend the standard (e.g., DITA base types).

Popular standards

There are many different standards available, this section describes but a few.

This section describes:

- DocBook
- DITA
- SCORM
- SPL

DocBook

DocBook is a very popular content model for software documentation. Most production-worthy XML authoring tools come with an implementation of DocBook as one of the standard sample DTDs.

It's a very broad model which can accommodate pretty much any kind of system software guide that you want to create. However, with over 300 individual elements, it is very complex and really demands a customization layer to filter out unwanted tags and structure. To mitigate that, the DocBook developers have designed the model and DTD to facilitate the use of a customization layer. This layer is built on top of DocBook to simplify the models actually used by authors.

If you're looking for a stable, software documentation model, look at DocBook; it's a mature, tested model. It has been used and tuned over a long period of time and over varied applications. As a result, it has built-in models to cover most, if not all, typical software documentation applications. It is complete! You could implement DocBook "straight out of the box."

One advantage of DocBook is that there is a plethora of stylesheets available to use with docbook for output. This includes stylesheets for simple HTML pages, web site HTML (with TOCs and indexes), MicroSoft HTML Help (.chm files), and PDF.

DocBook's biggest disadvantage, is despite the stylesheets for online formats, it remains a book-based, software documentation model.

For more information on DocBook, go to <http://www.oasis-open.org/home/index.php>

DITA

The Darwin Information Typing Architecture (DITA) is an XML-based architecture for authoring, producing, and delivering technical information with a focus on information interchange and reuse. The DITA architecture defines four layers:

- delivery context - the processing and delivery context
- typed topic structures - the formal content structure
- common structures - metadata and table structures that can be shared with any topic
- shared structures - content models for structures that can be used in all documentation

DITA is a powerful model that focuses on reuse with a topic-based core. A common misconception is that DITA, like DocBook, defines everything you could possibly want in content models. In fact, the DITA DTD defines only base models and its developers expect that you will create your own topic types to meet your own information needs. This means you must model your content to determine your topic types.

DITA is currently growing in leaps and bounds in terms of popularity. Like DocBook, the DITA package includes stylesheets to help users take content from authoring to delivery. XML vendors have started announcing specific support for DITA as the interest in the model has grown in the information development community.

This makes it an excellent possibility as the core of a reuse implementation.

For more information on DITA, go to <http://www-106.ibm.com/developer-works/xml/library/x-dita1/>

SCORM

SCORM (Shareable Content Object Reference Model) is an XML-based method for representing course structures. SCORM describes:

- how to create web-based learning content that can be delivered and tracked by learning management systems
- what a learning management system must do to properly deliver and track SCORM compliant learning content.

The SCORM model is not intended to replace instructional design models. It does not specify how tracking information is stored and what reports are generated, what pedagogical or training models should be used, or how learner information is assembled. Instead, it specifies how SCORM-compliant learning content is:

- delivered through a web browser
- described by metadata
- organized as a collection or sequence
- packaged in such a way that it can be imported by a compliant learning management system or into a repository used by such a system.

SCORM provides a model for the distribution of packaged learning objects for reuse; it is not a model for designing effective reusable learning objects. Selecting a SCORM-compliant tool will enable you to share your reusable learning objects, but you must still use sound pedagogical methods for designing learning models.

For more information on SCORM, go to <http://www.adlnet.org/>

SPL

The life sciences industries are having to adjust an increasing number of defined standards. One such standard is Structured Product Labeling (SPL). In these industries, labeling includes package inserts, prescribing information, product information, and product characteristics. The standard is based on XML and includes schemas to support the content structure. The goals of the standard includes:

- Improved access to information by medical professionals and consumers
- Faster communication of labeling information to improve risk management
- Improved consistency and interchange of the information.
- Increased opportunity to systematically process the information.

Key features of the standard include modularization of information and increase facility for reuse.

For more information on SPL, go to <http://www.fda.gov/oc/datacouncil/spl.html>

Others

There are many, many content standards and defined content structures available. You can find many more examples of published standards and structures at <http://xml.coverpages.org/xmlApplications.html>

Summary

There are many reasons to support standards in information design and creation. Sometimes you have to (for regulatory compliance or industry conformance), other times you are just taking advantage of opportunities (information interchange and technological opportunity).

Standards can impact your processes and content management strategy in different ways. Some standards can be approached as simple output formats, similar to HTML and PDF. You author in structures of your choice and transform the content into the standard's markup scheme at output. Other standards provide structures for the authoring process, and are used directly by the author to create modules of information.

The first step to understanding how to implement a standard structure and the impact on your organization of that implementation is to thoroughly understand the content that you create and how you create it. A thorough audit of your content and processes, along with content models, is necessary to understand how to implement standards that you might author in, how content can be reused in authoring and output, and how content must be transformed.

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