An XML-based information architecture for learning content, Part 1: A DITA specialization design

Use DITA XML to develop reusable learning content

John P. Hunt (john_hunt@us.ibm.com)
DITA Learning Architect
IBM

Robert Bernard (bbernard@us.ibm.com)
DB2 Training Developer
IBM

Can topic-based DITA XML provide the basis for developing an information architecture for single-sourced XML learning content? This article builds directly on the rich background about reusable content and e-learning delivery in the learning and training fields. Here in Part 1, the authors posit a set of extensions to DITA XML that provide the starting point for a unifying content model for learning. In Part 2, they test their assumptions against pilot content from a training course developed to support a component feature of IBM® DB2 Query Monitor™, and then report their findings and suggest important next steps.

The ability to have a standard definition for educational information is appealing because it allows you to:

- Minimize duplicate effort (reuse)
- Use material from alternate sources (repurpose)
- Supply course topics to alternate deliverables (repurpose)
- Build to company-wide standardized methods
- Create custom courses quickly

This article offers some background on reusable learning objects and e-learning, and then proposes a high-level design for a unifying content model for learning based on the DITA XML content standard. Part 2 of this series reports results of a pilot project to test the usefulness of this content design against actual course content.
Background

Several industry trends in technical communication and technical training have converged over the past several years, all with the goal of fostering and capitalizing on the value of reusable content. These include:

- Reusable learning objects (RLOs)
- The Sharable Content Object Reference Model (SCORM) standard for Web-based e-learning
- The Darwin Information Typing Architecture (DITA) standard for XML-based content

These emerged in response to several specific challenges faced by technical content providers in the 1990s. Shorter product development and delivery cycles, the need to support multiple output formats (no longer just books), the shift to online and Web-based content delivery, and the move to componentized products have all increased focus on the need for a content architecture that promotes greater reuse, repurposing, and integration of information assets, both within and across organizations.

Reusable learning objects

Reusable learning objects, or RLOs derive from the pioneering work of learning content designers at several companies, including Autodesk®, Oracle®, and Cisco®. According to author Peder Jacobsen, an RLO represents "a discrete reusable collection of content used to present and support a single learning objective" (see Resources). With RLOs, it is possible to gather a pool of information objects and make them available for reuse and repurposing in a variety of learning delivery contexts.

For example, Figure 1 shows how an RLO on using an electronic address book might provide an instructional unit for one learning module about using a messaging system, for another on setting a team schedule, and yet another on sending invites for an e-meeting.

Figure 1. A reusable learning object (RLO)

The business value of such RLOs stems directly from the benefits of such content reuse. For example, with RLOs, you can:

- Use existing content to create new courses or penetrate new markets
- Accommodate multiple delivery channels (Internet, intranet, print, and more)
- Streamline content revisions by updating discrete content
- Improve course development time and efficiency
• Assemble new courses and other deliverables from existing content, in whole or in part

Due to these virtues, learning objects have gained widespread appeal. However, while they suggest a general approach for developing reusable content, learning objects do not in themselves provide a standard way to package and deliver that content to users.

The SCORM standard for e-learning

The Shareable Content Object Reference Model (SCORM) emerged in response to this need for a standard packaging and delivery model for learning. Born from a U.S. Department of Defense initiative (see Resources), SCORM provides a suite of capabilities that enable interoperability, accessibility, and reusability of Web-based e-learning content.

SCORM builds directly on the RLO foundation, adopting the more general term sharable content object, or SCO. With SCORM, SCOs provide specific launchable assets that are available for use and reuse in multiple learning contexts and deliverables.

DITA XML: A unifying content reuse architecture for learning

While both learning objects and SCORM bring to the fore the need for sharable content, they both specifically leave open the question of a particular format or structure for this content. In fact, it can literally be said that SCORM is a packaging and delivery specification in search of a content model. This brings us to the third and most recent trend: DITA XML.

The Darwin Information Typing Architecture (DITA) provides an XML-based standard for creating and delivering content. Spawned from a workgroup effort at IBM and now an OASIS open standard, DITA has it roots in best practices for technical authoring.

It’s thus no surprise that key characteristics of DITA directly address the crucial building blocks for developing reusable learning objects in general, and SCORM sharable content objects in particular. These DITA reuse characteristics include topics, topic types, domains, maps, and specialization:

• A DITA topic forms the most basic information unit -- short enough to be easily readable, but long enough to make sense on its own.
• A DITA topic type defines the role of a topic within an information set.
• A DITA domain defines vocabularies for common use across more than one topic type.
• A DITA map applies context to the topics. With maps, you organize different combinations of topics for different outputs and deliverables.
• Finally, DITA specialization provides a mechanism for deriving new topic types, new domains, and new map types as extensions to existing domains or types.

Table 1 summarizes how DITA responds to several of the key learning reuse characteristics.

| Table 1. Learning reuse and DITA |
|----------------------|-----------------|
| For this learning reuse characteristic... | DITA provides... |

An XML-based information architecture for learning content, Part 1: A DITA specialization design
<table>
<thead>
<tr>
<th>Stores learning content as discrete, reusable information units, known as either reusable learning objects (RLOs) or sharable content objects (SCOs)</th>
<th>Content created and stored as topic-based information units, with maps to organize groups of topics into larger units or modules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual RLOs or SCOs have no specific context, making them available for reuse in multiple deliverables</td>
<td>Context-free topics that can stand on their own for use in map-driven output contexts</td>
</tr>
<tr>
<td>RLO or SCO constructs typed for learning</td>
<td>An extensible set of topic types to structure content for different purposes</td>
</tr>
<tr>
<td>SCORM specifies an XML manifest, which aggregates and sequences SCOs to form larger units of instruction, according to identified learning objectives</td>
<td>Maps that group topics by goals, tasks, or objectives and organize them for deliverables, either by sequence or by relationship to other topics</td>
</tr>
</tbody>
</table>

While RLOs in general and SCORM in particular assume a particular content model for learning, they do not pre-suppose a particular content format.

**Extend DITA to support learning content**

Core DITA provides the starting point to develop a content model for learning. However, learning content and delivery have specific needs that go beyond what's available with the core DITA topic types and processing model.

Fortunately, the DITA specialization architecture provides a built-in method to extend DITA to support new content needs associated with learning.

Specifically, we developed the following DITA extensions to support learning content:

- **New topic types** that support learning-specific topic types to provide lesson overviews and objectives, summaries, exercises, and assessment content.
- A new **content domain** to describe specific content vocabularies that are used across the DITA topic types needed to support learning. For example, **instructor notes** represents a content domain that's required in all of the learning topic types.
- A new **map domain** to organize collections of DITA learning topics for assembly and delivery as a learning course.
- A DITA **process model** that puts it all together for designing, writing, and delivering learning content.

**DITA topic types for learning**

All DITA topic types specialize from a top-level generic type. The new content types needed to support learning build on the core DITA topic types, and extend to a new main branch of the DITA topic hierarchy as shown in Figure 2.
Figure 2. DITA topic types for learning

The `learningBase` specialization
The learning types all specialize from a `learningBase` type, which provides common content structures for the other learning types. `learningBase`, in turn, specializes directly from the DITA generic base topic.

The `mainpoints` element in learning content
A key content element in `learningBase` is `mainpoints`. This content element emerged as a key need for learning content, and serves several purposes, depending on the deliverable:

- For instructor-led classroom training, `mainpoints` provides the content for display in instructor overheads.
- For an e-learning or SCORM deliverable, `mainpoints` provides high-level summary content about a topic.
- For a printed deliverable, such as instructor notes or student reference, `mainpoints` provides section introductions.

Including core concept, task, and reference content
The `learningTopic` type provides a container for `mainpoints` content, plus nested content from the core DITA concept, task, and reference topic types. This nested content can be incorporated in one of three ways:

- Authored directly in the `learningTopic` type
- Pulled in for reuse or repurposing from existing topics through the DITA content reference (conref) mechanism
- Included in the output stream through a map

Summary of the learning topic types
Table 2 summarizes the purpose and core content elements for each of the learning topic types.

<table>
<thead>
<tr>
<th>Learning type</th>
<th>Provides this information...</th>
<th>And uses these learning-specific elements...</th>
</tr>
</thead>
</table>
| `learningBase`      | Common content structures for the other learning types | • `mainpoints`  
                      |                                               | • duration  
<pre><code>                  |                                               | • description |
</code></pre>
<table>
<thead>
<tr>
<th>Learning Type</th>
<th>Description</th>
<th>Main Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning Overview</td>
<td>Introduces and identifies the learning objective for a course, module, or lesson</td>
<td>Objectives, Resources</td>
</tr>
<tr>
<td>Learning Topic</td>
<td>Task, reference, and conceptual information that provide procedures, background, and other details that support a learning objective</td>
<td>Mainpoints, Duration, Description, Objectives, Prereqs, Audience, Add-Info, Comments, Outline, Resources</td>
</tr>
<tr>
<td>Learning Summary</td>
<td>Recaps the learning objectives and key points of a course, module, or lesson</td>
<td>Mainpoints, Review, Nextsteps, Resources</td>
</tr>
<tr>
<td>Learning Practice</td>
<td>Labs and practice exercises that reinforce a learning objective and prepare for an assessment</td>
<td>Mainpoints, Duration, Description, Objectives, One or more sections</td>
</tr>
<tr>
<td>Learning Assessment</td>
<td>Graded items to measure progress against identified objectives of a lesson, module, or course</td>
<td>Mainpoints, One or more sections</td>
</tr>
</tbody>
</table>

**DITA content domain for learning**

The learning specialization design provides a content domain to identify specific kinds of learning vocabularies available for use within and across the learning types.

The initial learning content domain defines an `instructornote` element, which is based on the core DITA `footnote` element and provides a way to include instructor notes anywhere in the body of any learning topic.

As additional domain-specific vocabularies for learning are identified, you can add them to this learning-domain specialization.

**DITA map domain for learning**

A DITA map domain specifies a set of specialized `topicref` elements in a map, and can be used to define the design pattern for a particular map topic structure. For learning content, a map domain can formalize a map structure with a structured sequence of references to learning topic types. In this way, a map domain for learning instantiates the sequencing and grouping of DITA topics that comprise an RLO.

For example, a map domain can define a learning object as a specific sequence of overview, supporting task, concept, and reference topics, a summary, and optional practice and assessment topics, as follows:
A DITA process model for learning content

DITA supports an overall process model for designing, developing, and delivering content, which can be extended to support learning content.

Key phases in a DITA end-to-end process model for developing and delivering learning content include:

1. Identify and model learning objectives and goals
2. Organize objectives into lessons and modules
3. Identify existing topics and develop new topic-based content that supports these objectives
4. Develop topic content for labs, exercises, and assessments, as appropriate
5. Write overviews and summaries for each objective and the overall course
6. Structure the topics for delivery in a particular course with a map
7. Use XSLT to process the map and topics for the particular deliverable

Summary

Topic-based DITA XML provides essential ingredients for developing reusable learning content. The DITA specialization architecture enables you to develop new DITA topic types that support learning content. With DITA maps, you can define a design pattern that ties these topics together into an overall information architecture for learning content.

Read Part 2 of this series to see how the IBM DB2 team applied this design and the overall DITA process model for designing, developing, and delivering content to an actual DB2 training course. Part 2 will also include a download with the DITA specialization schemas and sample content files, for use with the DITA Open Toolkit.
Resources

- Read the next article in this series, *An XML-based information architecture for learning content, Part 2* and test the authors’ assumption against pilot content from a training course (developerWorks, August 2005).
- Learn more about DITA in these developerWorks articles:
  - Introduction to DITA (October 2003)
  - Specializing topic types in DITA (October 2003)
  - Design patterns for information architecture with DITA map domains (September 2004)
  - Specializing domains in DITA (October 2003)
  - Why use DITA to produce HTML deliverables? (October 2003)
  - The DITA FAQ (November 2004)
- Find additional information on DITA at the OASIS Cover Pages.
- Get more background on learning objects and SCORM:
  - Cisco Systems Reusable Information Object Strategy (Cisco Systems, 1999)
    - Downloadable PDF.
  - The Sharable Content Object Reference Model (SCORM) (Advanced Distributed Learning, 2005)
- Join the OASIS Darwin Information Typing Architecture (DITA) Technical Committee.
- Find hundreds more XML resources on the developerWorks XML zone.
- Get a copy of the DITA Open Toolkit (SourceForge.net, 2005).
About the authors

John P. Hunt

John Hunt charts the overall technical direction and strategy for product user assistance for IBM portal and collaboration software. He has designed award-winning help systems and spearheaded his team's migration to DITA XML and a topic-based information architecture. For DITA, he has driven the move to support learning content.

Robert Bernard

Bob Bernard is a certified training specialist and training course developer for IBM DB2 software. He is chief evangelist for encouraging IBM learning developers to make the move to structured authoring with DITA.

© Copyright IBM Corporation 2005
Trademarks
(www.ibm.com/developerworks/ibm/trademarks/)