SIGDOC

What are OASIS DITA topics?

Stanley Doherty, Ph.D.

Guidepost

Purpose: This resource provides a conceptual overview of DITA topics and topic types.

Learning objective(s):

- Understand what DITA means by a "topic."
- Review how DITA maps organize topics by reference.
- Review the concept of information typing.
- Review the DITA topic types available with OASIS DITA 1.3.
- Review the elements common to all topic types.
- Answer some of the common questions about topic-based authoring.

Prerequisites:

- Familiarity with basic concepts associated with structured content.
- Familiarity with XML elements and attributes.
- Familiarity with a DITA editor.

Revision history (ACM-016)

Date	Author	Revision summary
TBD	Stan Doherty	First public posting

DITA is often associated with the concept of "topic-based authoring." To unpack this concept, we need to address several questions:

What is a DITA topic?

From the first release of DITA in 2005, the notion of a "topic" has been central.

A topic is a unit of information with a title and content, short enough to be specific to a single subject or answer a single question, but long enough to make sense on its own and be authored as a unit. [Source]

Just to set a baseline, this ACM document is not topic-based and has been authored in a linear fashion. Its content is authored in one flow and is stored in one file. It is organized with multiple headings (sections), but each section is cemented in place and cannot be moved around easily.

By way of contrast, topic-based authoring sees a publication as a collection of multiple "chunks" of content, each stored in a separate file (`*.dita`). This is a topic. Each topic can be relocated easily within and between publications. In part, it is common sense.

Forty-four years before DITA, the United States Navy published some research on modular writing called the Quick Reading Comprehension (QRC) method. At the time, large numbers of scientists were contributing to large technical publications and were getting in one another's way having to author unwieldy chapters and sections. In addition, the readers of these technical publications were spending an unreasonable amount of time identifying the content that was relevant to their individual work. The QRC project leads proposed that the content in large publications be broken down into topics with headings familiar to both authors and readers. Authors work independently on topics for which they have expertise and leave the final placement and integration of the topics to the end.

> The individual-unit idea may readily be extended beyond the primary reader (scanner) aids listed above to sections of the main text. Here each section or subsection is confined to discussion of a single topic. There is no cross-referencing. This permits the sectional topics to be prepared at any appropriate time and in no particular order. They are done piecemeal.

Note: You can download all the DITA sources used in this article from GitHub:
- https://github.com/acm-sigdoc-structured/dita-topic-types

How do DITA maps organize topics into a publication?

A DITA map is an XML document (`*.ditamap`) that contains mostly references to DITA topics. When you are working in a DITA editor, you create your DITA topics and then reference them from your working map. Here's how these DITA topic files appear in a DITA editor when they are referenced from a map.
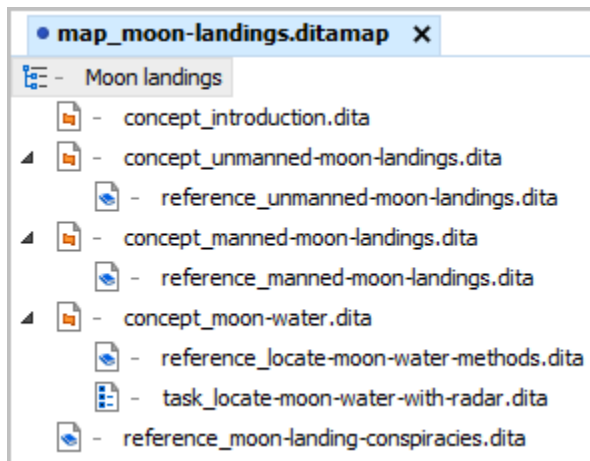
Most DITA editors can toggle between the display of filenames (above) and the display of topic titles (below).
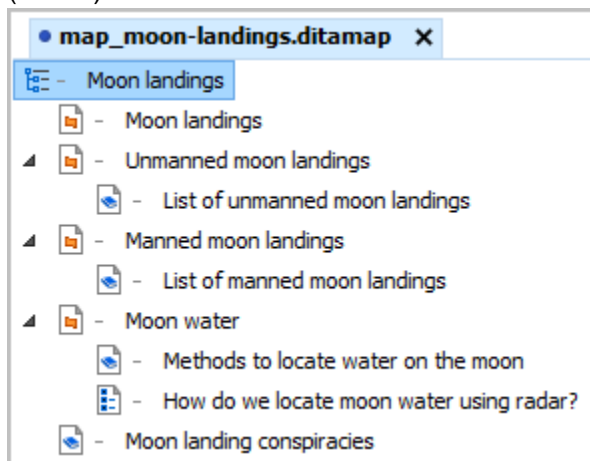
Each entry in this map is a topic reference (`<topicref>`) to a DITA topic file (`*.dita`). Here's how the underlying XML markup looks.

```xml
<map>
    <title>Moon landings</title>
    <topicref href="concept_introduction.dita"/>
    <topicref href="concept_unmanned-moon-landings.dita">
        <topicref href="reference_unmanned-moon-landings.dita"/>
    </topicref>
    <topicref href="concept_manned-moon-landings.dita">
        <topicref href="reference_manned-moon-landings.dita"/>
    </topicref>
    <topicref href="concept_moon-water.dita">
        <topicref href="reference_locate-moon-water-methods.dita"/>
        <topicref href="task_locate-moon-water-with-radar.dita"/>
    </topicref>
    <topicref href="reference_moon-landing-conspiracies.dita"/>
</map>
```

If I want to relocate or promote/demote a particular topic, I simply move the entry in the map.

DITA processors read these entries in the DITA map and generate the desired PDF, HTML, or webhelp output. Here's how the generated PDF table of contents looks.

Contents

It's fair to say that DITA topics provide the content for a publication, while DITA maps provide the organization.

How does information typing affect topic-based authoring?

Forty years before DITA, Robert Horner proposed that most content in most technical publications could be categorized into one of six "information types": procedures, processes, principles, concepts, structures (description), or facts. Horn called it [information mapping](). He argued that readability and comprehension are closely related to the consistency of information types within a document. The more frequently a reader is forced to switch between different information types, the more frequently the reader needs to reset expectations about the content. Instead of a document making 50 switches, Horner argued that we ought to minimize the number of switches by aggregating similar information into specific types of topics. When readers encounter a procedure topic or reference topic, they know what to expect within the boundaries of that topic type.

Some of the [benefits]() of using information types when authoring are:

- *Author focus*: Writers developing content in a particular information type are more likely to restrict content appropriate for that information type.

- *Reader focus*: Readers of content in a specific information type will have a more coherent reading experience.

- *Content economy*: In the process of developing content for a specific information type, far less extraneous or tangential content will creep into the writing. Technical cruft often does not make the cut.

- *Reuse planning*: With a clearer picture of the types of information in a doc set, it becomes easier to identify opportunities for reusing conceptual, reference, or task content.

In addition to adopting topics as the unit of authoring, DITA embraced information typing. Here are the most frequently used topic types. They ship with OASIS DITA 1.3 and are bundled with all DITA editors and component content management systems.

Topic type	Description
Concept	Concepts provide background that helps readers understand essential information about a product, a task, a process, or any other conceptual or descriptive information. Concept topics address the questions, "What is this?" or "How does it work?"
Reference	Reference topics provide data that supports users as they perform a task. Reference topics might provide lists and tables that include product specifications, parts lists, and other data that are often "looked up" rather than memorized. A reference topic can also describe commands in a programming language or required tools for a series of maintenance tasks. Reference topics address the question, "What details do I need to know?" or "What are the parts of this thing?"
Task	Tasks are the essential building blocks to provide procedural information. A task information type answers the "How do I?" question by providing precise step-by-step instructions detailing the requirements that must be fulfilled, the actions that must be performed, and the order in which the actions must be performed. The <task> topic includes sections for describing the context, prerequisites, expected results, and other aspects of a task.
Generic	Generic topics are untyped. The generic topic type should only be used if authors are not trained in information typing or when a specialized topic type is inappropriate.

Our sample publication on moon landings contains four concept topics, four reference topics, and one task topic. DITA editors often display small icons next to each topic reference in DITA maps to identify different topic types.
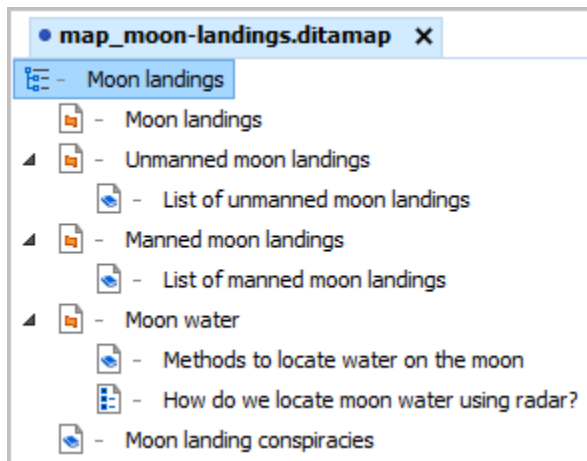
How do I create different topic types?

DITA editors and component content management systems make it easy to create instances of specific topic types. To create an instance, I simply choose **File - New** and select the appropriate topic type.
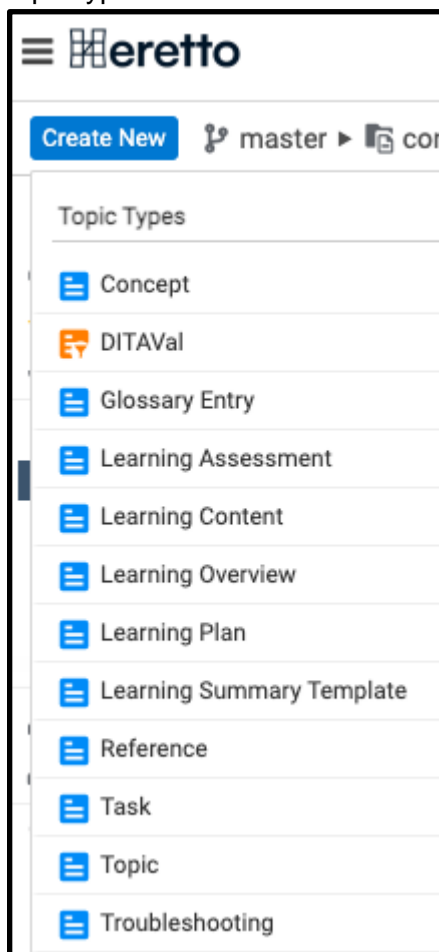
If I were to choose to create a reference topic, for example, the DITA editor or CCMS would read the XML document type definition (DTD) named `reference.dtd` and make a copy (instance) available to me to name and save. The XML markup in that new topic instance declares that it is a reference topic and is derived from the `reference.dtd` document type definition.

```
<!DOCTYPE reference PUBLIC "-//OASIS//DTD DITA Reference//EN" "reference.dtd">
```

Each topic type has a unique DTD from which authors create instances.

What are common and type-specific elements for topic types?

You might expect that the elements in each topic type would be completely unique, but they are not.

All the popular topic types can contain the following elements:

- `<prolog>`: Provides a container for metadata relevant to the topic, for example `<audience>`, `<author>`, `<component>`, `<keywords>`, `<prodinfo>`, and `<resourceid>`. Processors and search optimization tools use this metadata to create links to this topic from other topics sharing relevant metadata.

- `<title>`: Provides the title of the current topic. This title appears in generated output as topic or section headings, TOC entries, and cross-reference hypertext.

- `<shortdesc>`: Provides a brief description of topic content. This short description helps the reader answer the question, "Is this topic relevant to what I am looking for?"

- Common elements in running text: `<p>`, `<section>`, `<ol>`, `<ul>`, `<simpletable>`, `<example>`, or `<image>`.

In a DITA editor, here is how a concept topic might look with these common elements.

```
concept  title  Moon landings  title

shortdesc  Short Description: A moon landing refers to the event when a spacecraft
successfully lands on the surface of the Moon.  shortdesc

prolog  author  Author:        Commander Joan Gilroy  author
resourceid  Resource Identifier: AppId: NASA-311        AppName: NASA-App
resourceid  prolog

conbody  p  The most famous moon landing occurred on July 20, 1969, when NASA's
Apollo 11 mission, commanded by Neil Armstrong, successfully landed on the Moon.
Armstrong and astronaut Buzz Aldrin became the first humans to walk on the lunar surface,
while Michael Collins remained in orbit around the Moon in the command module.  p

p  This article explores the most prominent landings (manned and unmanned) and one of
the most important motivations for landings – the quest for water.  p

note  ✎  ⌄  Note:

p  This content was sourced from GenAI.  p  note  conbody  concept
```

Certain other elements are relevant to only one topic type:

- *Reference topics*: `<resyn>` (for syntax diagrams) and `<properties>` (for syntax arguments).

- *Task topics*: `<context>`, `<prereq>`, `<steps>`, `<step>`, and `<result>`. Formal task topics are perhaps the most restrictive in terms of which elements you can put where, but they offer many benefits in terms of readability. Here is what a task topic might look like in a DITA editor.

task title **How do we locate moon water using radar?** title

shortdesc Short Description: Ground-penetrating radar (GPR) is a powerful tool for detecting water on the Moon. shortdesc

taskbody prereq **Before you begin:** ▷*Pre-requisite*◁: You need to send and receive radar radio waves from an orbiting spacecraft. prereq

context **About this task:**

p ▷*Context*◁: GPR works by sending radio waves into the lunar surface and analyzing the reflected signals. p context

steps

1. step ▷The spacecraft transmitter sends radio radar waves to the lunar surface. ◁

   info These waves penetrate the surface to varying depths, depending on the material they encounter. info step

. . .

4. step ▷The analysis systems on the spacecraft transmit the raw data and the working map back to Earth. ◁ step

steps

result **Results:**

p ▷While radar is a valuable tool, it's often used in conjunction with other instruments, such as neutron spectrometers and infrared spectrometers, to confirm the presence of water. p result

What are some common concerns about topic-based authoring?

For authors who have had little or no experience with topic-based authoring, the transition can be challenging. Everyone learns at a different pace and in different ways, so there is no global success formula beyond getting solid training and sharing your experiences as you go. That said, many writers making the transition have some legitimate concerns.

- If I am not sure what the best topic type might be for something I am writing, what do I do?

  Writers do not need to make a decision about using a particular topic type when they are in the early stages of writing. I often draft content in a simple text editor when I am figuring things out. If I want to draft the content in DITA markup without making a decision about a topic type, I work in the DITA generic topic type. It is the least restrictive topic type and supports my moving draft

content into more appropriate concept, reference, or task topic types.

- Can I split one large topic into multiple, smaller topics?

  Yes - as writers gain experience authoring and mapping topics, the average size of DITA topics tends to get smaller. Some concept and reference topics contain `<section>` elements that are logical candidates to copy-paste into a new topic. Some tools automate this process of segmenting sections into separate topics.
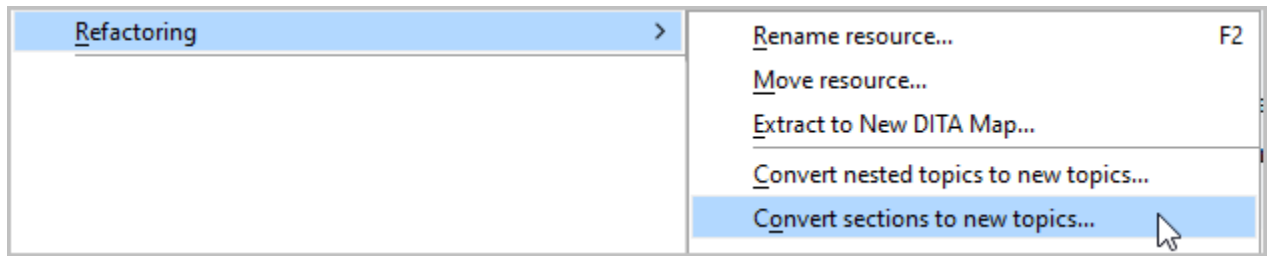
- Can a DITA editor convert the markup in one topic type to another topic type?

 Yes - popular DITA editors can automate the conversion of markup.
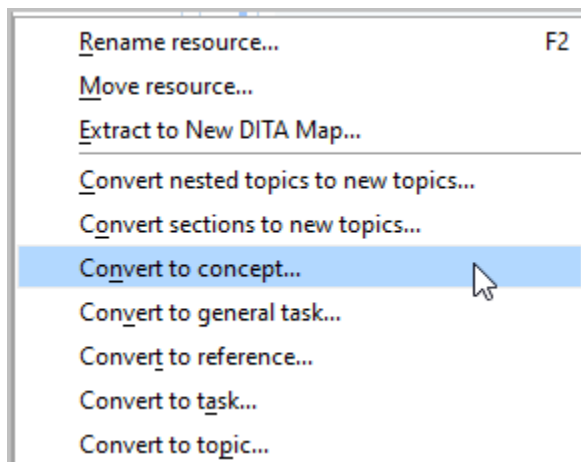
- Will I need to revise some of the content when I move content from one topic type to another?

  Yes – beyond converting the underlying markup, you'll need to rework some of the writing to support robust concept, reference, and task topic structures.
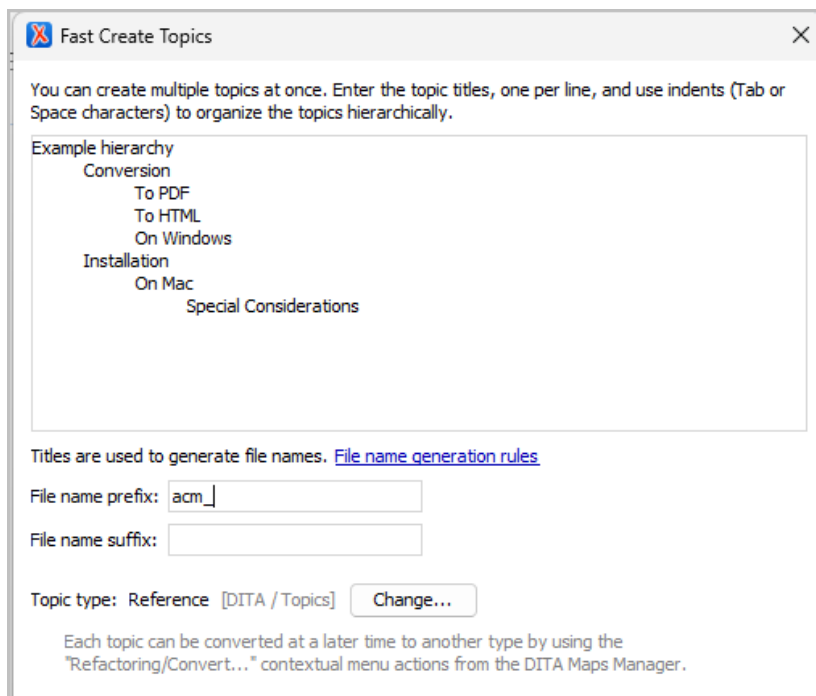
- Can I import unstructured content into a DITA topic?
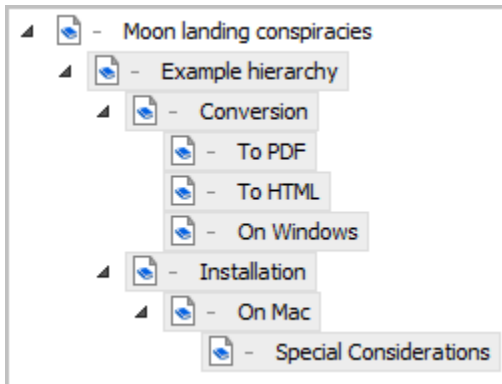
  Yes - DITA editors and component content management systems provide conversion tools.
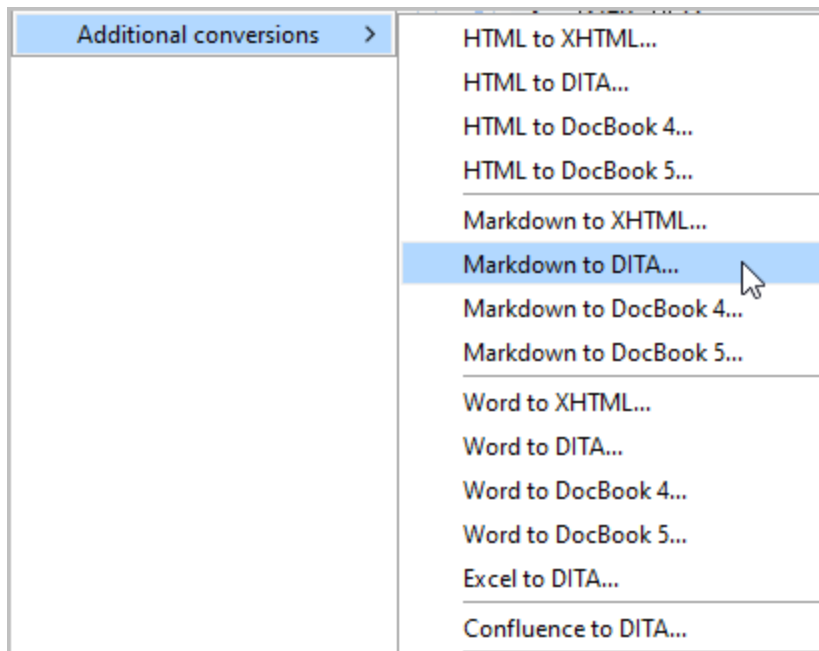
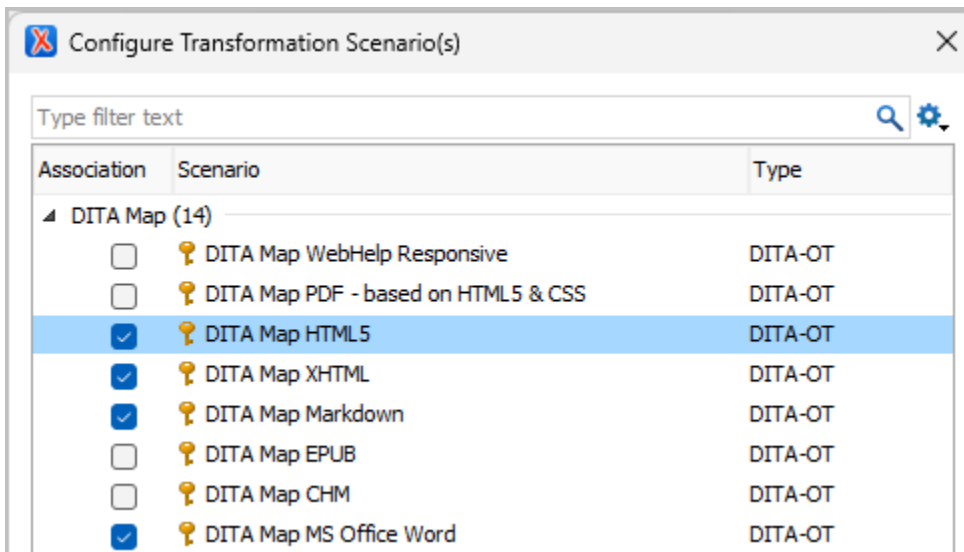- Can I create multiple DITA topics from an outline?

  Yes - some editors can convert a plain-text outline into a set of topics and then add them to your current map.

- Can I export multiple DITA topics to a single MS Word, HTML, or Markdown file?

  Yes - the DITA Open Toolkit provides multiple transformations from modular (DITA) to linear document formats.

Acknowledgements

- Scot Marvin - editorial review
- JoAnn Hackos - technical review
- Carlos Evia - instructional review

Thank You!

See https://acm-sigdoc-structured.org to learn more about committee activities, available resources, and volunteer opportunities.

Created by members of the ACM SIGDOC Committee on Structured Authoring and Content Management