# Publication: Review H: Chunking (00813611-DD\_1)

## Topic: Example: Using chunk to split documents (DA00715317)

### Paragraph-level comments

Similarly, because setting <code>chunk="split"</code> on the map element sets a default for the entire map, the following change to the original map would result in *every* referenced DITA document being split into one document per topic. The only source document not affected by this split is about.dita, because it only contained a single topic to begin with.

Annotation	Reviewer	Status	Туре	Date	Topic version
In the sample code that follows, please make the <b>chunk="split"</b> text bold.	gjoseph	updated	comment	23/1/2022 11:48:08	
Done. Marking this comment <b>COMPLETED</b> .	keberlein	updated	comment	25/1/2022 14:09:56	

## **Topic: Processing chunk="combine" (DA00715322)**

### **Topic-level comments**

processor specific.

Annotation	Reviewer	Status	Type	Date	Topic version
Somewhere, we should capture the effects of chunk="combine" on topic <prolog> metadata. Specifically, all <prolog> metadata in subordinate "chunked" topics gets removed from generated output. Probably nothing that we can do about it, but users should be surprised when all their beautiful metadata disappears. :-)</prolog></prolog>		updated	comment	19/1/2022 11:29:12	
@Stan, a couple of thoughts here:	keberlein	updated	comment	20/1/2022 14:57:35	
<ul> <li>I suspect that this is an implementation detail that will depend on how a given processor chooses to implement support for chunking. @Robert?</li> <li>I don't know how else a processor COULD handle this. It would not make sense to combine all the metadata How a processor would handle this is probably best left up to the implementation.</li> <li>This point is probably material for a user guide for a particular implementation, especially if how metadata is handled during a chunk="combine" operation is</li> </ul>					

----

Discussed at TC call on 25 January 2022. Adding a draft comment to the archSpec content about metadata, stating that we might (or might not) want to make a statement about most rendering and processing for metadata is by necessity implementation-specific.

Marking this comment COMPLETED.

We could make a clear statement that what a processor does with the metadata is implementation-specific ... Thoguhts?

### Paragraph-level comments

Setting chunk="combine" instructs a processor to combine the reference source documents for rendering purposes. A single result document is generated.

Annotation	Reviewer	Status	Туре	Date	Topic version
Setting chunk="combine" instructs a processor to combine the reference d source documents for rendering purposes. A single result document is generated.		updated	change	19/1/2022 11:20:41	
Done. Marking this comment <b>COMPLETED</b> .	keberlein	updated	comment	20/1/2022 12:42:48	

Setting chunk="combine" instructs a processor to combine the reference source documents for rendering purposes. A single result document is generated.

Annotation	Reviewer	Status	Type	Date	Topic version
Style question for the spec. Do we want to use attribute="value" or setting @attribute to value?  I can't quite decide if it's a good shorthand or not. I'm not against it, just wanted to ponder it.	zlawson	updated	comment	23/1/2022 02:11:14	
I think the most precise wording is "Specifying chunk="combine"" And even better if we can state "Specifying foo="bar" on the fancyWidget element	keberlein	updated	comment	25/1/2022 15:37:20	
Here we can't quite do that because one can specify @chunk on map and topicref elements					
@Robert?					
Marking this comment CLOSED.					

The following rules apply:

Annotation	Reviewer	Status	Type	Date	Topic version
Use the "syntactic that" within all of the list items below. For example (emphasis added), "When specified on the root element of a map, all source DITA documents <b>that are</b> referenced by the map are treated as one DITA document."					
This is important for translation and ESL readers. Somewhere in the TC document repository, there is a PDF about this. It's an article authored by John Kohl, who wrote <i>Global English</i> .	keberlein	updated	comment	20/1/2022 12:49:28	
<b> </b>					
Done. Marking this comment <b>COMPLETED</b> .					

• Once chunk="combine" is specified on a map, branch, or map reference, all source DITA documents grouped by that reference are treated as a single resource. Any additional @chunk attributes on elements within the hierarchy are ignored.

Annotation	Reviewer	Status	Type	Date	Topic version
Once chunk="combine" is specified on a map, branch, or map reference, all source DITA documents within the scope of that reference are treated as a single resource. Any additional @chunk attributes on elements within the hierarchy are ignored.	sdoherty	updated	change	19/1/2022 11:21:43	
@Stan, we don't want to use the term "scope" unless we are referring to a key scope. Changed to read as follows: "When chunk="combine" is specified on a map, map branch, or map reference, all source DITA documents that are grouped by the reference are treated as a single resource. Any additional @chunk attributes on elements within the grouping are ignored."	keberlein	updated	comment	20/1/2022 12:30:55	
Marked as COMPLETED.					

# Topic: Example: Using chunk to split nested documents (DA00715331)

### Paragraph-level comments

Special attention is necessary when evaluating the map hierarchy that results from splitting a documents with their own nested topics.

Annotation	Reviewer	Status	Type	Date	Topic version
Special attention is necessary when evaluating the map hierarchy that results from splitting a documents with their own nested topics.	gjoseph	updated o	change	23/1/2022 11:50:29	2
Changed to read "Special attention is necessary when evaluating the map hierarchy that results from splitting documents that contain nested topics."  Marking this comment <b>COMPLETED</b> .	keberlein	updated o	comment	25/1/2022 14:12:48	2

• middle.dita has nested topics, so results in its own new hierarchy within the map. Content from the nested topic reference is now located within the reference to the root topic from middle.dita, but after any references to child topics.

Annotation	Reviewer	Status	Туре	Date	Topic version
middle.dita has nested topics, which so results in its own new hierarchy within the map. Content from the nested topic reference is now located within the reference to the root topic from middle.dita, but after any references to child topics.	gjoseph	updated	change	23/1/2022 11:56:00	
Done. Marking this comment <b>COMPLETED.</b>	keberlein	updated	comment	25/1/2022 14:14:29	

# **Topic: Example: Combining topics within a split context** (DA00715356)

### Paragraph-level comments

Assume the following map, where chunk="split" on the root element means that all topic documents within this map structure are split by default, but a branch within the map sets chunk="combine".

	Annotation	Reviewer	Status	Type	Date	Topic version
	e sample that follows, please bold the two es of the chunk attribute to make it easy for the dentify.	gjoseph	updated	comment	23/1/2022 12:16:27	
Done. N	Marking this comment COMPLETED.	keberlein	updated	comment	25/1/2022 14:17:31	

# **Topic: Processing chunk="split" (DA00715357)**

### **Topic-level comments**

Annotation	Reviewer	Status	Туре	Date	Topic version
If the split was on a topicref for a single topic with sections, could it split the sections as individual docs?	dstevens	updated	comment	19/1/2022 20:29:55	
Quite simply, no. Chunking only applies to topics.  Marking this comment CLOSED.	keberlein	updated	comment	19/1/2022 22:45:25	
I would recommend that we explicitly state split doesn't work with sections. I honestly didn't realize that. I avoid multiple topics in a file, so I misunderstood how split could work. While I realize it's probably listed somewhere, I didn't realize that the glossentry was based off of topic. To me it felt like a section (because I was working in a topic set up by someone else, so I never bothered to read the spec about how it specifically worked.)  Again, this could be me wishing for more user content in the spec, or just clarifying that split refers to dita topics (because I didn't see or really understand the distinction on first reading).	zlawson	updated	comment	31/1/2022 01:34:05	
After discussion, Robert and I added the following paragraph to the "About the @chunk attribute" topic:  "The @chunk attribute only operates on topics and nested topics. It does not operate on other topic content, such as sections."  Marking this comment as <b>COMPLETED</b> .	keberlein	updated	comment	1/2/2022 12:18:00	

### Paragraph-level comments

• When specified on a <topicref> element that refers to a source DITA document, it indicates that all topics within the referenced document should be rendered as individual documents.

Annotation	Reviewer	Status	Type	Date	Topic version
Let's start by saying I've never used chunk except to combine. Now if we want to split something, it needs to be a single topic with multiple topics in it. If it was actually a map and referenced by a mapref or a topicref with a format of ditamap,	dstevens	updated	comment	19/1/2022 20:30:38	

it would be automatically split, correct? If so, this explains why I've never used it because I instruct my clients never to combine topics into a single dita topic, but to use maps instead. So you would never have a <mapref chunk="split"> because it already does that? But you might have a &lt;mapref chunk="combine">. No action needed, just looking to ensure I understand it.

The use cases that I've run into for using @chunk to "split" (I'm not going to list the DITA 1.3 token for @chunk, because I'd have to look it up) are the following:

- Legacy content that uses the ditabase topic type (might originally been built to generate only PDF)
- Content that was migrated to ditabase as an intermediate format, but which has not made its way through the entire conversion process
- Companies that have used the glossgroup topic type, for the convenience of authors, but want to generate single HTML topics for each glossentry topic

keberlein updated comment  $\begin{array}{c} 20/1/2022\\ 02:34:19 \end{array}$ 

I hope others will chime in with the use cases that they are familiar with.

-----

Discussed at TC call on 25 January 2022. The other use case mentioned was the old IBM message specialization.

Marking this comment CLOSED.

• When specified on an element such as <topicgroup> that does not refer to a topic or result in a published topic, the attribute has no meaning.

Annotation	Reviewer	Status	Туре	Date	Topic version
In this list item, we refer to a "topic"; all the other list items refer to "source DITA documents." Is the usage here correct?	keberlein	updated	comment	20/1/2022 19:36:51	
Changed to read "When specified on an element such asthat does not reference a source DITA document or result in published output, the attribute has no meaning."  Marking this comment <b>COMPLETED.</b>	keberlein	updated	comment	25/1/2022 14:44:30	

• When specified on the root element of a map, chunk="split" sets a default operation for all source DITA documents in the map (outside the context of relationship tables). The default split value is used except where a combine value is encountered, in which case combine takes over for that entire branch.

Annotation	Reviewer	Status	Type	Date	Topic version
Suggest replacing the parenthetical with the following change (emphasis added here for clarity)" " sets a default operation for all source DITA documents in the <b>navigation structure of</b> the map."	keberlein	updated	comment	20/1/2022 19:38:28	
Done. Marking this comment <b>COMPLETED</b> .	keberlein	updated	comment	25/1/2022 14:45:26	

# Topic: Example: Using chunk to render a single document from one branch (DA00715338)

#### Paragraph-level comments

```
input.ditamap:
   <map>
   <title>Lesson plan</title>
   <topicref href="goals.dita">
   <!-- more goal topics -->
   </topicref>
   <topicref href="firstLesson.dita">
   <!-- more tasks in the first lesson -->
   </topicref>
   <topicref href="nextLesson.dita">
   <!-- more tasks in the next lesson -->
   </topicref>
   <!-- More branches -->
   </map>
  firstLesson.dita:
   <task id="firstLesson">
   <title>Starting to work with scissors</title>
   <shortdesc>This lesson will teach ...</shortdesc>
   <taskbody><!-- ... --></taskbody>
   </task>
  nextLesson.dita:
   <task id="nextLesson">
   <title>Advanced cutting</title>
   <shortdesc>This lesson will introduce complicated shapes...</shortdesc>
   <taskbody><!-- ... --></taskbody>
   </task>
```

```
Annotation Reviewer Status Type Date Topic version

input.ditamap: Lesson plan firstLesson.dita:
Starting to work with scissors This lesson will teach ...

nextLesson.dita: Advanced cutting This lesson will introduce complex complicated shapes...

gjoseph updated change 23/1/2022 11:30:51
```

14:21:12

When output requirements demand that portions of the map be combined into a single document, adding chunk="combine" to a branch of the map instructs a processor to render one document that combines all topics in that branch. This is particularly useful when the topics need to be rendered independently for other contexts, or when the way topics are contributed makes creating a single source document impossible.

Annotation	Reviewer	Status	Type	Date	Topic version
When output requirements demand that portions of the map be combined into a single document, adding chunk="combine" to a branch of the map instructs a processor to render one document that combines all topics in that branch. This is particularly useful when the topics need to be rendered independently for other contexts or when the way topics are contributed makes creating a single source document impossible.	sdoherty	updated	change	19/1/2022 11:24:44	
It took me a LONG TIME to figure that your change was removing a comma!  Removed the comma, also removed the adjective "particularly". Marking this comment <b>COMPLETED</b> .	keberlein	updated	comment	20/1/2022 14:27:33	

In the following sample, the original map is updated with @chunk attributes to indicate that each lesson branch is rendered as a single result document; topics in the first branch with goals.dita are not affected as a result of the @chunk attribute.

Annotation	Reviewer	Status	Type	Date	Topic version
In the following sample, the original map is updated with @chunk attributes to indicate that each lesson branch is rendered as a single result document, † T topics in the first branch with goals.dita are not affected as a result of the @chunk attribute.	gjoseph	updated o	change	23/1/2022 11:32:45	
Done. Marking this comment <b>COMPLETED.</b>	keberlein	updated o	comment	25/1/2022 14:23:23	

**Topic** Reviewer Status Type **Annotation** Date version input.ditamap: Lesson plan firstLesson.dita: Starting to work with scissors This lesson will teach ... 23/1/2022 gjoseph updated change 11:34:34 nextLesson.dita: Advanced cutting This lesson will introduce complex complicated shapes... keberlein updated comment Done. Marking this comment **COMPLETED**.

# Topic: Example: Using chunk to combine all documents into one (DA00715335)

### **Topic-level comments**

Annotation	Reviewer	Status	Type	Date	Topic version
In figure 1, in file goals.dita, change the tile from "gals" to "goals". This fix must be made to Figure 3 as well.  The change mode of ditaweb looses all the elements, so I felt a comment was more effective here	gjoseph 1	updated	comment	23/1/2022 11:27:23	
Done. Marking this comment <b>COMPLETED.</b>	keberlein	updated	comment	25/1/2022 14:26:31	

A couple of comments on this topic:

keberlein updated comment 27/1/2022 19:06:43

- The titles of the figures need to be parallel
- I'm a little concerned about the figures that contain snippets from BOTH the maps and topics. The content of codeblocks needs to be valid DITA ...
- If the code blocks contain file names, they should be commented out.
- If we DON'T need file names, then let's not use them. For example, I think we don't need to indicate that the root map is named input.ditamap.

I'm going to rework this topic and then send it to you.

----

Reworked the example topic in consultation with Robert. This comments applies to all the examples, so making this comment as **ACCEPTED**.

### Paragraph-level comments

The result of evaluating this @chunk attribute is equivalent to the following map and topic document; content from all topics within the map is combined into a single result, with a topic order and topic nesting structure that match the original map hierarchy.

Annotation	Reviewer	Status	Type	Date	Topic version
The result of evaluating this @chunk attribute is equivalent to the following map and topic document. Content from all topics within the map is combined into a single result, with a topic order and topic nesting structure that match the original map hierarchy.	sdoherty	updated o	change	19/1/2022 11:23:47	
Changed to read as follows (emphasis added): " The content from all topics within the map is combined into a single result document, with a topic order and topic nesting structure that match the original map hierarchy."  Marking this comment COMPLETED.	keberlein	updated (	comment	20/1/2022 14:31:33	2

# Topic: Example: Using chunk to combine a group of topics (DA00715316)

#### Paragraph-level comments

The @chunk attribute on <topichead> also results in a single DITA document. In many applications, a <topichead> is equivalent to a single title-only topic; in that case, the chunked result is equivalent to a root topic

with the title "Heading for a branch", containing as child topics the content of both inhead1.dita and inhead2.dita. If <topichead> is ignorable in the current processing context, the chunked result would be equivalent to processing <topicgroup> (a single DITA document with the content of both inhead1.dita and inhead2.dita).

	Annotation	Reviewer	Status	Type	Date	Topic version
tit to as in th D	ne @chunk attribute on also results in a single DITA ocument. In many applications, a is equivalent to a single le-only topic. I in that case, the chunked result is equivalent a root topic with the title "Heading for a branch", containing child topics the content of both inhead1.dita and head2.dita. If is ignorable in the current processing context, e chunked result would be equivalent to processing (a single ITA document with the content of both inhead1.dita and head2.dita).		updated	change	19/1/2022 11:25:54	
	Done. Marked as <b>COMPLETED.</b>	keberlein	updated	comment	20/1/2022 14:35:40	

## Topic: Chunking (DA00509128)

### Paragraph-level comments

Content often needs to be delivered in a different granularity than it is authored. The @chunk attribute enables map authors to specify that multiple source documents are combined into a single document for delivery, or that a single source document is split into multiple documents for delivery.

Annotation	Reviewer	Status	Туре	Date	Topic version
Content often needs to be delivered in a different granularity than it is authored. The @chunk attribute enables map authors to specify that multiple source documents be combined into a single document for delivery or that a single source document be split into multiple documents for delivery.	sdoherty	updated	l change	19/1/2022 11:15:42	
Changed to read "Content often needs to be delivered in a different granularity than it is authored. The @chunk attribute enables map authors to specify that multiple source documents should be combined into a single document for delivery or that a single source documentshould be split into multiple documents for delivery."	,	updated	l comment	20/1/2022 12:39:15	
While we normally avoid the term "should" where we can, I think it is very appropriate and should be used here. Also, the recent reading of OASIS boilerplate indicates that they are now making a distinction between the normative words					

(MUST, MAY, SHOULD, etc.) and where they are used in lower-case.

Marking this comment COMPLETED.

### **Topic: About the chunk attribute (DA00715339)**

#### **Topic-level comments**

**Topic Annotation Reviewer Status** Date **Type** version The page is not responsive . . . cannot enter changes. Paragraph-1: Using "best" implies some receommended or "best" practice. Perhaps "appropriate" would be more neutral. sdoherty updated comment  $\frac{19/1/2022}{11:20:21}$ chunk="combine": Consider "serialize" instead of the last 'combine". chunk="split": "each single source XML document" -- are not 'each" and "single" redundant here? 1. Changed "best" to "optimal". 2. Let as-is; I do not think that "serialize" is the correct word choice here. The term has distinct meanings in the both the general computer science communities, keberlein updated comment and I don't think that either meaning is correct here. @Robert? 3. Removed the word "single" Marking this comment **COMPLETED**. dstevens updated comment 19/1/2022 The fact that processors may create their own naming scheme 20:23:18 has caused my clients quite a bit of grief because that naming scheme is typically smething like "chunk12345" and that

The fact that processors may create their own naming scheme has caused my clients quite a bit of grief because that naming scheme is typically smething like "chunk12345" and that number changes every time the content is processed, meaning there is no predictable file name to reference in a link. We have some customers therefore using @copy-to to provide a predictable name. But that attribute is going away, correct? So what is the recommendation for this situation? Obviously my experience is all on the combine side, so being able to specify the appropriate file name for the single output could be done with a single attribute. I don't know how the split situaiton would work, which is likely why there is no way to specify the names. But I'm still curious about recommendations.

I left in my ramblings, but once I got to the last example, I

realized this issue is addressed. Could we add something to the bullet here that authors can specify a predictable file name using keys as shown in the last example?

Discussed in the TC call on 25 January 2022. Marking this comment as CLOSED.

25/1/2022 keberlein updated comment

"The chunk attribute does not cascade." But if it is specified on a <map> it does, if I'm reading the topics correctly. When specified on <map> the action applies to all topics within the map (split or combine) and on a branch it applies to dstevens updated comment 21:03:52 all referenced topics in the branch (maybe only combine -- it's only written that way in processing chunk='combine"). So this seems inconsistent to me.

Your comment made me go off and look at 5.3.1 "Cascading of metadata attributes in a DITA map" in the current spec draft. The shortdesc reads as follows:

"Certain map-level attributes cascade throughout a map, which facilitates attribute and metadata management. When attributes *cascade*, they apply to the elements that are children of the element where the attributes were specified. Cascading applies to a containment hierarchy, as opposed to a element-type hierarchy."

keberlein updated comment  $\frac{20/1/2022}{15:10:08}$ 

We need to improve the definition of cascade!

But to answer your comment, @chunk does not cascade. The presence of @chunk is a signal to a processor to do something with the topics referenced by the element on which the @chunk attribute is specified.

Marking this comment as **ACCEPTED**.

If we are going to use the terms "source document" and "result document," we need to do so consistently. This is a global comment for this entire review.

keberlein updated comment

Marking this comment ACCEPTED.

gjoseph updated comment 23/1/2022

10:35:27

Regarding this setence: Other tokens can be defined by applications, but support for those tokens will vary.

We discuss the ability to extend processing via custom tokens further down in "Using the chunk attribute for other

purposes". Therefore I think we should either delete this sentence completely or reduce it to something like: Applications can define additional tokens.

This topic has been restructured significantly. Marking this comment as **CLOSED**.

keberlein updated comment  $\frac{25/1/2}{17:42}$ :

### Paragraph-level comments

## Topic: Using the chunk attribute for other purposes (DA00715326)

### **Topic-level comments**

Annotation	Reviewer	Status	Туре	Date	Topic version
If we add this section here, wouldn't we need to add it every attribute that could be specialized. What makes @chunk special?	sdoherty	updated	comment	19/1/2022 11:22:52	
Agree with Stan. It seems this is here to acknowledge that we removed some tokens from 1.3.	dstevens	updated	comment	19/1/2022 21:04:37	
We state this because applications are permitted to add their own custom, implementation-specific, tokens for @chunk. That's not the case for most DITA attributes that take spec-defined tokens. The other @attribute (might be others as well) that falls in this category is @cascade. For @cascade, we explicitly state in a normative statement that processor MAY use their own tokens  FYI, @chunk is not a specialized attribute. The only specialized attributes that we ship in the base are the standard, conditional-processing attributes.  Marking this comment CLOSED.		updated	comment	19/1/2022 22:42:58	

### Paragraph-level comments

### Topic: Example: Managing links when chunking (DA00715334)

### **Topic-level comments**

Annotation Reviewer Status Type Date Topic version

The statement that it is "straightforward" made me laugh; I dstevens updated comment 19/1/2022 can think of no client who would agree with this statement. I 20:58:32 think that the first part of this intro before the semicolon should just be deleted. It is an opinion statement. While Dawn's clients are not the audience for the specification, I agree that the short description needs to be rewritten, so that it better captures what the example topic is trying to communicate. The content is not straight-forward keberlein updated comment  $\frac{20/1/2022}{14:47:22}$ and simple ... Marking this comment as **ACCEPTED**. In the short description, remove the semi-colon and make it two sentences. keberlein updated comment  $\frac{20/1/2022}{14:37:55}$ Done. Marking this comment **COMPLETED**. Assume that the map above is a root map or is used by another map that does not otherwise render the three topic documents. gjoseph updated comment  $\frac{23/1/2022}{12:28:12}$ In that case, the following is true: keberlein updated comment  $\frac{25/1/2022}{14:31:28}$ Done. Marking this comment **COMPLETED**. Please remove the semicolons from the last paragraph: There is no way to unambiguously link to the child document that will result from splitting splitThis.dita. This is because it is only possible for the element using @chunk to associate a key definition with the first or root topic in the document. While other key definition elements can be used to associate keys with other topics in the same document, that can only be updated comment 23/1/2022 13:28:49 done outside of the navigation context that uses @chunk; as a gjoseph result, a processor cannot guarantee whether the intended link target is the split topic from the @chunk context, or a use of the same topic in the second context. It is possible for an implementation to define its own way to resolve this ambiguity; however, if a situation requires both multiple instances of split topics and unambiguous crossimplementation links to those split topics, alternate reuse mechanisms need to be considered. keberlein updated comment  $\frac{25/1/2022}{14:33:22}$ Done. Marking this comment **COMPLETED.** 

# **Topic: Example: When chunk is ignored (DA00715343)**

### **Topic-level comments**

Annotation	Reviewer	Status	Туре	Date	Topic version	
ditaweb is not recognizing block elements on this topic, so I have to use a comment instead :(						
For this sentence: The @chunk attribute on the is ignored; it does not cascade, and there is no referenced topic, so it has no effect.						
Rewrite as something like: The @chunk attribute on the is ignored. The attribute value does not cascade and there is no referenced topic, so the attribute has no effect.	gjoseph	1-4- 1		23/1/2022		
Regarding the last sentence in the second bullet item, please break up the sentence around the semicolon: If the <topichead> is treated as a title-only topic, it cannot be split further; if it is ignored for the current processing context, it is no different than the <topicgroup>.</topicgroup></topichead>		gjosepn	updated	comment	12:04:23	
Replace with: If the <topichead> is treated as a title-only topic, it cannot be split further. If it is ignored for the current processing context, it is no different than the <topicgroup></topicgroup></topichead>						
Done. Marking this comment <b>COMPLETED.</b>	keberlein	updated	comment	25/1/2022 14:39:14		

### **Paragraph-level comments**