Simplifying Information Types in DITA: Expanding the Foundation for Enterprise Applications

This document describes a proposal by the DITA for Enterprise Business Documents Subcommittee to define abstract versions of the DITA base topic types intermediary between <topic> and the DITA 1.2 <concept>, <task>, and <reference> topic types. These abstract information types would provide more inclusive models than the existing DITA 1.2 topic types. The new information types would be semantically ancestral to the broadest range possible of less abstract business topic types used in specific disciplines. Examples of these disciplines include technical documentation, research and development documentation, and policies and procedures.
Simplifying Information Types in DITA
Expanding the Foundation for Enterprise Applications

CONTENTS
Introduction......................................................................................................................... 1
The Opportunity: Respond to Enterprise Adoption of DITA ............................................. 1
Rising to the Challenge ............................................................................................... 3
The Proposal: Create a layer of abstract specializations ............................................... 4
Create new topic types to represent abstract information types ..................................... 4
Remove domain specializations ...................................................................................... 5
Remove Select attributes ............................................................................................... 5
Create <procedural>, an abstract task topic ................................................................. 5
Create <descriptive>, an abstract reference topic .......................................................... 5
Create <explanatory>, an abstract concept topic ........................................................... 6
Inheritance ....................................................................................................................... 6
Extension of abstract topic types in business documents ................................................... 6
Impact on implementation ............................................................................................. 10
Conclusions and Implications for the DITA Technical Committee ................................. 11

Introduction
This document describes a proposal by the DITA for Enterprise Business Documents Subcommittee to define abstract versions of the DITA base topic types intermediary between <topic> and the DITA 1.2 <concept>, <task>, and <reference> topic types. These abstract information types would provide more inclusive models than the existing DITA 1.2 topic types. The new information types would be semantically ancestral to the broadest range possible of less abstract business topic types used in specific disciplines. Examples of these disciplines include technical documentation, research and development documentation, and policies and procedures.

The Opportunity: Respond to Enterprise Adoption of DITA
A growing number of organizations are drawn to DITA specifically because of enterprise orientation and promise of a more interoperable XML infrastructure across departments and applications. These organizations understand the historical focus of technical publications within the DITA community. They also recognize the broader architectural foundation within the specification, and are looking to the DITA Technical Committee to aid in the creation of an enterprise model of information types. An important part of this enterprise model is a more inclusive set of topic types that are semantically and
ancestrally related both to the current technical publication topic types, and to yet undefined topic types that will be created for other functional areas outside of technical publications. The need for a semantic relationship is based upon the very broad applicability of concept, task, and reference to most written business content. The need for an ancestral relationship results from a belief that if properly conceptualized, these more inclusive forms of the three primary topic types will accommodate both the technical publication specializations and the yet undefined specializations for a broad range of enterprise documents.

The members of the DITA Enterprise Business Documents Subcommittee are of the view that now is the time to address these enterprise topic requirements—before organizations adopt a myriad of home-grown models without the benefit of guidance. While some of these models might work well, we are concerned that the majority will fall short of what can be achieved with the support of the experienced DITA community, and in falling-short will reflect poorly on the standard itself.

We also view this as the right time to address this need because the work invested by the Technical Committee in DITA 1.2, while unrelated to these particular enterprise requirements, has provided a much needed foundation for a more delineated model that will help organizations see how each department or application fits into the overall architecture. We find that the requirements for enterprise topic types are associated both with usability for authors, and enterprise fitness for information architects.

As would be expected authors of business topics report that DITA presents them with choices for markup that are unrelated to their business requirements. In particular, they report frustration with the number of elements included through the programming, software, and user interface domain specializations, which they find difficult to understand without a great deal of training and for which they have little or no use.

To a lesser degree, business authors find the same lack of relevance with some of the elements defined within the <task> and <reference> topic specializations. For example, the <refsyn> element defined in the <reference> topic specialization has little application in the general business environment and its purpose is not obvious from its name. Other little-used elements from the <task> topic specialization include <stepxmp>, <tutorialinfo>, and <syntaxdiagram>. In working with business users outside of the technical publications department, we find that elements that are not related to requirements do not just add bulk—they add a significant amount of confusion that directly impacts user adoption.

Information architects have the need for an enterprise level of abstraction for the <task> and <reference> content models. As might be expected, attempting to derive widely varying business document specialization from the existing technical publishing oriented content models has led to considerable frustration. For example, the <reference> topic type is a natural topic type to use for requirements topics, such as those found in common business documents, like requests for proposal (RFP) or functional specifications. The body content of such topics often starts with paragraphs providing details about the requirement. However, the existing <reference> topic type does not allow <p> elements at the start of the <refbody> element, so it cannot be used for such topics. Similarly, an information architect cannot specialize the <reference> topic to allow <p> elements in
<refbody>: this would relax restrictions on the content model, which is not permitted by DITA’s specialization rules. The only recourse for the information architect would be to specialize a new <requirement> topic type from the <topic> base topic type.

Similarly, information architects find that the <task> topic type content model is unnecessarily complex for a very common type of business topic called the standard operating procedure. A standard operating procedure consists primarily of one or more instructions, each of which rarely requires more than a handful of the elements currently defined in the <task> topic specialization. The constraints mechanism proposed for DITA 1.2 would certainly enable an information architect to define a simplified <task> topic type for standard operating procedures, but the constraints would be quite complex, since they would have to explicitly exclude many elements, attributes, and domain specializations from the existing <task> topic specialization. Again, the architect’s best recourse might be to go back to the <topic> base topic type and specialize an <sop> topic type for standard operating procedures.

Persisting with this strategy would promote a great proliferation of new topic types specialized from <topic>, making DITA seem ever more unwieldy and complex. The truth is that DITA should, and can, support a very clean taxonomy of information.

Rising to the Challenge

The subcommittee believes the current situation is the natural outcome of the initial focus on technical publications by the architects of DITA. While it is clear that a fuller range of enterprise business documents is enabled in the initial DITA architecture, the initial efforts had a more specific focus. As a result, the existing <task> and <reference> topic specializations, and the existing software, programming, and user interface domain specializations which were originally intended for complex software documentation applications are too subject-matter specific. The Enterprise Business Documents Subcommittee was chartered to explore the full range of enterprise document types as appropriate applications for DITA.

Much of the recent activity of the DITA Technical Committee and its various subcommittees actually stems from this original focus on technical publications which resulted in the incorporation of specializations of the DITA architecture. This persisted in DITA 1.1 and in the drafts for DITA 1.2:

- Development of the constraints mechanism
- Loosening of the <task> topic specialization to accommodate the requirements of machine industry documentation
- Creation of the Adoption Subcommittee to ease these barriers to adoption
- Creation of the Technical Documentation Subcommittee to separate the development of the existing <concept>, <task>, and <reference> topic specializations from other specializations and from the base topic type

Upon first examination, it might seem that the packaging conventions being implemented in DITA 1.2 would fully support our desire to see an evolution of the enterprise characteristics of DITA. However, we are concerned that the segregation of the various topic specializations from each other and from the <topic> base topic type will prevent them from sharing common semantics and common structures efficiently.
DITA has been designed for interoperability and reuse. As we extend the content model to the enterprise it becomes very important to maintain this level of interoperability to take advantage of enterprise content reuse (e.g., across marketing, sales, tech pubs, and training). If the topic specializations for enterprise business documents are segregated by deriving them all directly from `<topic>`, none of the semantic elements that have already been developed, particularly the specializations in the DITA 1.1 `<task>` and `<reference>` topic types, will be available for reuse, either by the subcommittee or subsequently by users. This and subsequent specializations will create even further segregation of content and increase the incidence of content silos rather than bridge silos. Providing an abstract layer will promote enterprise content sharing and reuse.

The Proposal: Create a layer of abstract specializations

The subcommittee proposes that we establish a layer of information type specializations of the `<topic>` base topic type, from which the existing topic and domain specializations can in turn be specialized. This intermediate layer of specializations would provide simpler content models using elements and attributes with names that are more immediately understandable to most English-speaking authors. By creating a new intermediate layer of root elements, we ensure backwards compatibility for existing DITA content.

The subcommittee believes the essential semantic functions of the `<concept>`, `<task>` and `<reference>` topic types have great value in application to enterprise business documents where:

- A `<concept>` topic type’s main function is to explain
- A `<reference>` topic type’s main function is to describe
- A `<task>` topic type’s main function is to instruct

These functions meet the requirements of many enterprise business document types.

- A product brochure, for example, consists primarily of a topic that explains the product’s value proposition and at least one topic that describes the product’s features.
- A request for proposal (RFP) consists primarily of a topic that explains the need a proposal must address, at least one topic that describes the requirements the proposal must meet, and at least one topic that instructs responders how to submit the proposal.

To better move the DITA standard towards an enterprise orientation, the subcommittee therefore proposes for DITA 1.3 to provide abstractions of the existing `<concept>`, `<task>`, and `<reference>` topic types defined in DITA 1.2.

Create new topic types to represent abstract information types

We propose to name the root elements with adjectives based upon the function of each information type. By using an adjective for the root element (procedural rather than procedural_information or procedure), we discourage content creation using the abstract information types while supporting the creation of specialized elements named with nouns. For example:
• <task> represents *procedural information*, therefore label the root element <procedural>
• <concept> represents *explanatory information*, therefore label the root element <explanatory>
• <reference> represents *descriptive information*, therefore label the root element <descriptive>

This will make it easier for architects outside of the technical documentation community to create meaningful specializations that will interoperate with those for technical documents.

Remove domain specializations

Remove all domain specializations from <topic>, <procedural>, <explanatory>, and <descriptive> topics. In particular, the *programming, software, and user interface* domains are problematic outside of the technical documentation community. This will reduce the number of elements and attributes for which there is little use in enterprise business applications. It is easy to include these domains in specializations for which they are useful.

Remove Select attributes

Remove *base, platform, product, audience, and otherprops* attributes from the simplified topic types. They can easily be added back in to community specializations for technical documents and business documents using attribute specializations from the *props* attribute. The removal of technical documentation select attributes will make it easier for new specializations to be developed outside of the technical documentation community.

Create <procedural>, an abstract task topic

This proposed change is similar to Feature #12011 for a simplified form of <task> topic type. The changes that we propose go further to simplify the <task> type.

• Start from a new root element describing the information type, e.g., <procedural>.
• Change the name of the <task> topic <cmd> element to <action>. This name is more understandable to non-technical authors. Specialize task/cmd from <action>.
• Add a <process> element as an optional child of the <taskbody> element, a sibling of the existing <steps> and <steps-unordered> elements. The <process> element broadens the applicability of the simplified <task> topic type to business applications. Specialize <process> from <section> rather than from <ol>, the ancestor of <steps>, so that a process can be represented with flexibility, e.g. as an image or as a semantically rich description that can be rendered as SVG or other rich media.

Create <descriptive>, an abstract reference topic

Simplify the <reference> topic type content model:

• Remove the <refsyn> element.
• Allow para-type elements in the <refbody> element (e.g. <p>, <note>, <fig>), as is currently allowed in the <concept> topic <conbody> element.
• Replace the <properties> element with a semantically-similar, yet looser specialization from <simpletable> so that it is more suitable as a container for
descriptive content called `<characteristics>`. The new model would allow for multiple value elements to describe each characteristic. The proposed content model becomes:

- characteristics
  { characteristichead \((optional)\) then characteristic \(one \ or \ more\) \}
- characteristichead
  { characteristiccolname \(one \ or \ more\) \}
- characteristic
  { characteristictype \(optional\), characteristicvalue \(any\), characteristicdesc \(optional\) \}

Create `<explanatory>`, an abstract concept topic
Introduce a new root element for the information type, e.g., `<explanatory>`. The subcommittee does not propose making any other changes to the existing `<concept>` or `<topic>` types, other than the aforementioned removal of the domain specializations and select attributes.

Inheritance
In the specialization inheritance hierarchy, the simplified topic types resulting from these changes will be intermediate between `<topic>` and the `<concept>`, `<task>`, and `<reference>` topic types of DITA 1.2. The changes are therefore backward compatible, since `<concept>`, `<task>`, and `<reference>` would be specialized from them. In addition, other business topic types, which cannot be specialized from `<concept>`, `<task>`, or `<reference>`, could be specialized from the new abstract topic types.

Extension of abstract topic types in business documents
The subcommittee has been studying enterprise business documents over the past two years and has concluded that the task-based DITA information model is defined too narrowly to reach its full potential for enterprise business document users. The three DITA information archetypes are well founded in the history of user assistance documentation pioneered by IBM and Lotus in the early 1990’s\(^1\). They were never intended to encompass all information types within an enterprise.

Several models have been examined as a starting point for developing an information model for business documents. The subcommittee is currently studying the Information Management Model\(^2\) as a metamodel for business documents and will be publishing whitepaper on the metamodel in the fall of 2009. This metamodel consists of 11 information types that include concept, task, and reference.

On further examination, we discovered that there were common semantic sub-structures across almost half of the 11 business information types. To effectively build these information types, we stepped back to analyze exactly how the information is conveyed.

---

\(^{1}\) Doyle, Bob et.al. (2009) [http://dita.xml.org/book/history-of-dita#1.0 “History of DITA”], DITA XML.org, OASIS

This exercise led to the creation of six basic abstract information types:

- Procedural
- Explanatory
- Descriptive
- Directive
- Criterial
- Temporal

These information types are not intended to be used directly as topic types, but rather serve as a basis for structural specializations. Table 1: Abstract information types describes the six abstract types, their key semantic substructures, and a sample of topic types that can be derived from them.

<table>
<thead>
<tr>
<th>Info Type</th>
<th>Description</th>
<th>Semantic Substructure</th>
<th>Creates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explanatory</td>
<td>Explains an idea or concept. It may define how something works, how something is made, what something is. Ancestor to &lt;concept&gt; topic, among others.</td>
<td><code>&lt;explanatory&gt;</code>&lt;br&gt;<code>&lt;title&gt;</code>&lt;br&gt;<code>&lt;shortdesc&gt;</code>&lt;br&gt;<code>&lt;expbody&gt;</code>&lt;br&gt;<code>&lt;example&gt;</code>&lt;br&gt;<code>&lt;related-links&gt;</code></td>
<td>Concept&lt;br&gt;Design&lt;br&gt;GlossEntry</td>
</tr>
<tr>
<td>Descriptive</td>
<td>Describes specific information about an object. It may describe characteristics of an object, a person, or a place. It may include pictures or illustrations of the item at a given point in time. Ancestor to &lt;reference&gt; topic, among others.</td>
<td><code>&lt;descriptive&gt;</code>&lt;br&gt;<code>&lt;title&gt;</code>&lt;br&gt;<code>&lt;shortdesc&gt;</code>&lt;br&gt;<code>&lt;descbody&gt;</code>&lt;br&gt;<code>&lt;detail&gt;</code>&lt;br&gt;<code>&lt;sublist&gt;</code>&lt;br&gt;<code>&lt;characteristics&gt;</code>&lt;br&gt;<code>&lt;related-links&gt;</code></td>
<td>Reference&lt;br&gt;Role&lt;br&gt;Resource&lt;br&gt;Location&lt;br&gt;Function&lt;br&gt;Feature</td>
</tr>
<tr>
<td>Procedural</td>
<td>Information used to instruct the reader on the steps required to perform a task. Ancestor to &lt;task&gt; topic, among others.</td>
<td><code>&lt;procedural&gt;</code>&lt;br&gt;<code>&lt;title&gt;</code>&lt;br&gt;<code>&lt;shortdesc&gt;</code>&lt;br&gt;<code>&lt;taskbody&gt;</code>&lt;br&gt;<code>&lt;prereq&gt;</code>&lt;br&gt;<code>&lt;context&gt;</code>&lt;br&gt;<code>&lt;section&gt;</code>&lt;br&gt;<code>&lt;process&gt;</code>&lt;br&gt;<code>&lt;steps&gt;</code>&lt;br&gt;<code>&lt;result&gt;</code>&lt;br&gt;<code>&lt;example&gt;</code>&lt;br&gt;<code>&lt;postreq&gt;</code>&lt;br&gt;<code>&lt;related-links&gt;</code></td>
<td>Task&lt;br&gt;Process&lt;br&gt;Procedure&lt;br&gt;Instruction&lt;br&gt;TestCase</td>
</tr>
<tr>
<td>Info Type</td>
<td>Description</td>
<td>Semantic Substructure</td>
<td>Creates</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Directive</td>
<td>Information used to direct or limit performance of a procedure. The statement is backed up with information to help the user understand the consequences and resolution of performance relative to the statement.</td>
<td><code>&lt;directive&gt;</code>&lt;br&gt;<code>&lt;title&gt;</code>&lt;br&gt;<code>&lt;statement&gt;</code>&lt;br&gt;<code>&lt;dirbody&gt;</code>&lt;br&gt;<code>&lt;detail&gt;</code>&lt;br&gt;<code>&lt;sublist&gt;</code>&lt;br&gt;<code>&lt;conditions&gt;</code>&lt;br&gt;<code>&lt;exclusions&gt;</code>&lt;br&gt;<code>&lt;penalties&gt;</code>&lt;br&gt;<code>&lt;remedies&gt;</code>&lt;br&gt;<code>&lt;related-links&gt;</code></td>
<td>Policy&lt;br&gt;Regulation&lt;br&gt;Guideline&lt;br&gt;Hazard&lt;br&gt;FAQ</td>
</tr>
<tr>
<td>Criterial</td>
<td>Information used to specify criteria that are measurable or comparable to other pieces of information. Allows for comparison of what was needed versus what was delivered.</td>
<td><code>&lt;criterial&gt;</code>&lt;br&gt;<code>&lt;title&gt;</code>&lt;br&gt;<code>&lt;shortdesc&gt;</code>&lt;br&gt;<code>&lt;critbody&gt;</code>&lt;br&gt;<code>&lt;detail&gt;</code>&lt;br&gt;<code>&lt;sublist&gt;</code>&lt;br&gt;<code>&lt;constraints&gt;</code>&lt;br&gt;<code>&lt;criteria&gt;</code>&lt;br&gt;<code>&lt;related-links&gt;</code></td>
<td>Objective&lt;br&gt;Requirement&lt;br&gt;Ability&lt;br&gt;Activity</td>
</tr>
<tr>
<td>Temporal</td>
<td>Information that reports on an event. Records specific time and/or place of the occurrence and summarizes the outcome.</td>
<td><code>&lt;temporal&gt;</code>&lt;br&gt;<code>&lt;title&gt;</code>&lt;br&gt;<code>&lt;summary&gt;</code>&lt;br&gt;<code>&lt;stamp&gt;</code>&lt;br&gt;<code>&lt;date-time&gt;</code>&lt;br&gt;<code>&lt;location&gt;</code>&lt;br&gt;<code>&lt;occurrence&gt;</code>&lt;br&gt;<code>&lt;temporalbody&gt;</code>&lt;br&gt;<code>&lt;details&gt;</code>&lt;br&gt;<code>&lt;sublist&gt;</code>&lt;br&gt;<code>&lt;related-links&gt;</code></td>
<td>Event&lt;br&gt;Incident&lt;br&gt;Problem&lt;br&gt;Report</td>
</tr>
</tbody>
</table>

Figure 1 Topic specializations showing abstract layer illustrates the separation of the abstract layer from the DITA 1.1 topic types and the business document topic types. The implication of this separation is that ownership is also separated allowing for custody of the abstract information types to reside with the technical committee while technical publications subcommittee and the enterprise business documents subcommittee maintain control over their respective topic definitions. This model does not preclude the creation of other abstract types or direct specializations from topic for other communities of practice.
Development of these new topic types is guided by the metamodel which might be thought of as an ontology of information types. This metamodel encompasses the well-known technical documentation requirements as well as other business documentation requirements that the Enterprise Business Documents Subcommittee has been examining.

The metamodel is organized by the fundamental questions Who, What, How, and Why. This is represented, without supporting discussion, in Figure 2 Enterprise Business Documents Information Metamodel. Error! Reference source not found.

This metamodel is sketched here only to show the larger context of the intermediate layer of abstract information types which is the topic of discussion in this paper. In our view, this or any alternative metamodel that might be adopted must provide for this layer.
Impact on implementation

The impact of these changes represents a fundamental change to the context and representation of the currently accepted DITA topic types. The most significant impact of these proposed changes will be to the educational aspect of the work of the Technical Committee, updating basic conceptual information about the DITA content model.

The actual technical implementation should involve only very straightforward changes to the DITA DTD/Schema and the OT publishing processes to deprecate formatted output to the new simplified topic types for cases where content cannot be generalized directly to <concept>, <task>, or <reference>. Rather than implementing constraints on the simplified <task> to create the DITA 1.1 <task> structure, as proposed in Feature #12011, the current <concept>, <task>, and <reference> types will require specialization from these proposed simplified information types.

All DITA 1.0 and 1.1 content should be fully backwards compatible with this proposed change.

Note that the introduction of the abstract layer should make it easier for packaging in DITA 1.3. If possible, the TC should consider how DITA 1.2 is packaged to support the introduction of abstract information types into the basic package for DITA 1.3.
Conclusions and Implications for the DITA Technical Committee

The subcommittee’s charge has been to identify and analyze classes of content broadly identified as “enterprise business documents” and harmonize these with the DITA standard.

The content and use cases driving DITA 1.0 and 1.1 were predominantly for technical user assistance documentation. Members and observers of the DITA Technical Committee have been bringing additional new use case domains into the picture. Education, Machine tools, pharmaceutical, and publishing interests are only a beginning.

The technical committee has a forest-and-trees challenge. As DITA 1.2 extends beyond this original base, the technical committee has begun to deal with the positioning of technical documentation relative to other use case domains. This is especially challenging because the usual technical committee process is to resolve particular technical problems at a fairly narrow level of detail in the workings of DITA, always with one eye out for architectural consequences. In the course of our work, the subcommittee members have come to recognize an urgent need for the technical committee to step even farther back and take in a very large perspective on the future prospects of DITA before making decisions that could trouble us in the future.

One of our earlier realizations as a subcommittee was that technical documentation is a special case of enterprise business documentation, but we were reluctant to push this view because of how very presumptuous it seems. Taking a larger view, we now see that DITA stands poised to grow into the most widely applied XML standard for language-borne information. We realize that as DITA broadens to handle web content, government content, and other information types, business documents will become a specialized case of more generalized narrative communication, and establishing new levels within the taxonomy are part of that process.

It is imperative that we clarify the foundations for specializations and constraint domains in such a way that our familiar <concept>, <task>, and <reference> topic types can each share common ancestry with the kindred topic types that will (not may) be developed for these wider use case domains. If we wait until advocates for these use cases develop their new base topic types from <topic>, it will be too late, and the future of DITA will be locked into the foreseeable awkwardness of information silos with consequent difficulties for adoption of DITA and sharing of DITA content which are still avoidable, if we act now.