

Topic Maps for Managing Classification Guidance

Vinh Lê
U. S. Department of Energy

James David Mason
Y-12 National Security Complex

Abstract

The Department of Energy (DOE) and its predecessor agencies have had a long history of developing and handling sensitive, classified information. Since the Manhattan Project, identification of classified information has been supported by principles and rules, called "classification topics," in guidance documents that are published for use by authorized classifiers across the DOE complex. Managing DOE classification guidance adds to the usual problems of document management. Coordination of multiple dependencies among master and derived documents and classification topics is needed to ensure consistency. The Topic Maps technique is being applied to organize classification guidance topics according to unique subjects so that duplicate topics, inconsistent topics, and gaps in classification guidance become obvious for corrective actions. Once topics are logically organized and linked, then when a change in a guidance topic is proposed, system users will know which related topics need a review or change. Development of an overall guidance-management system also involves creation of a new XML-based publishing system to replace the diverse and often inadequate tools used in the past. A document-management system to support the publishing system will require integration with the guidance-management system in addition to the conventional tools for revision control and file management. We have begun to construct the topic maps for guidance management and are in the process of refining a design of topic maps to manage the publishing process.



Topic Maps for Managing Classification Guidance

Table of Contents

1 The DOE Classification Guidance System.....	1
2 The Guidance Streamlining Initiative.....	2
2.1 Infrastructure Development Objective.....	2
2.2 Guidance Streamlining Objective.....	2
3 Central Publishing System Requirements.....	2
3.1 Document Structure.....	3
3.2 Coordination.....	3
3.3 Versioning.....	3
3.4 XML Publishing System Design.....	4
4 GSI and a Design for Managing the Content of Classification Guides.....	4
5 Topic Map for GSI.....	5
5.1 Ontology for Guidance Management.....	5
5.1.1 Objects.....	6
5.1.1.1 Guides.....	6
5.1.1.2 Guidance Topics.....	6
5.1.1.3 Keywords.....	7
5.1.1.4 Keystones.....	7
5.1.1.5 Coordination and Reference Items.....	7
5.1.1.6 Properties Associated with Guidance Topics.....	7
5.1.1.7 Source Documents.....	8
5.1.2 Association Roles and Links.....	8
5.1.2.1 Links around Guidance Topics.....	8
5.1.2.2 Links Involving Topics with Keystones and Keywords.....	10
5.2 Instance Pointers.....	13
5.2.1 Guides.....	13
5.2.2 Instance Guidance Topics.....	13
6 Using the Map.....	14
7 Future Directions for Guidance Management.....	15
Acknowledgements.....	16
Bibliography.....	17
The Authors.....	17



Topic Maps for Managing Classification Guidance

Vinh Lê and James David Mason

§ 1 The DOE Classification Guidance System

Since the beginning of the Manhattan Project, research and development activities in DOE have been distributed across many sites. Facilities were originally dispersed across the nation because of wartime conditions. Much of the basic research activities are not sensitive and are unclassified. However, applications for nuclear weapons designs, development, testing, and manufacturing are sensitive and often classified at secret level. Although most technologies related to the production and enrichment of nuclear materials are well known, DOE protects certain advanced technologies and unique design information.

Sensitive information products such as design drawings, assembly procedures, technologies, or design concepts are classified. The DOE office of Information Classification and Control Policy (ICCP) develops classification policy and provides guidance for determining what information is or is not classified in accordance with the Atomic Energy Act and applicable Executive Orders. The DOE classification guidance is documented in classification guides and is provided in the form of topics. Each classification topic describes the information (or rules, policy) and has a classification level (e.g., unclassified, confidential, secret, top secret). Classification guides are written for subject matter experts and for classification specialists who review information and make classification determination derivatively based on the original classification decisions as provided in classification topics. These authorized derivative classifiers (ADCs) are trained in specific technical subject areas and authorized by DOE to make derivative classification decisions.

Classification topics are often hierarchical. Frequently, rules appear in if-then-else patterns that reflect different combinations of conditions surrounding a piece of information. The master documents, sometimes referred to as "DOE headquarters classification guides," tend to offer general rules, such as that the association of certain classes of materials with particular products reveals a sensitive, unique concept and therefore is classified and protected. Derived documents, sometimes referred to as "local classification guides," make such rules specifically applicable to the specific environment or conditions of a DOE site. In local classification guides, each rule must be traceable to one or more rules in the headquarters classification guides, so that the logical hierarchy of classification guidance transcends individual documents.

The number of these classification topics has increased tremendously as we have gained additional knowledge in research and development, not to mention the changes in information protection schemes and strategies due to declassification. Declassification is a policy determination process of changing a piece of classified information so that it becomes unclassified. Overall classification policy does not change over time; however, declassification of individual pieces of information changes the information-protection strategies and schemes. The lack of an effective electronic document management system and a classified network constrains effective guidance management and as a result, creates opportunities for redundant and inconsistent classification policy and guidance. Classification topics and knowledge information became unmanageable in this *paper-based* system. Inefficient processes and ad hoc publishing rules and procedures have arisen to resolve these problematic symptoms, adding complexity to the real problems.

An ASCII-based classification guides system was developed a decade ago to solve some problems. However, this system was simply a limited collection of classification guides with search capability using keywords. Document reviewers use this system to search pertinent classification guidance for making classification determination. Guidance writers use this system to seek out classification

guidance that requires an update. Currently, a word processor (e.g., WordPerfect) is still being used as authoring and publishing tools. Approved guidance in a proprietary format is then converted to ASCII for input in this classification guides system. This guidance development and production process is tedious, inefficient, and prone to errors.

§ 2 The Guidance Streamlining Initiative

The classification Guidance Streamlining Initiative (GSI) was designed to solve these problems. The goal of this initiative is eventually to develop an integrated classification guidance system for users to efficiently write classification guidance and continuously update classification guidance for accuracy, clarity, and consistency. Two primary objectives were set towards meeting this goal. One is to develop and acquire an information management infrastructure that allows for the automation and facilitation of the classification guidance development, updates, publication, and dissemination. The other is to use this infrastructure and to-be-developed tools to streamline and improve classification guidance accuracy and consistency.

2.1 Infrastructure Development Objective

A needs analysis was performed and specific requirements were identified and specified in the *User Requirements for the Classification Guidance Database and Publishing System* [DOE 2001], prepared in early 2001. ISOGEN/DataChannel was selected to perform an analysis of the requirements and provided recommendations and a conceptual design of such a system [ISOGEN 2001]. Most notable among recommendations was the culture change needed in guidance authoring. In this new environment, guidance writers must learn an XML authoring tool to create content in a more structured way. This was the most challenging task. A scaled-down version of the system requirements was designed to evaluate the author's acceptance of the new structured-content authoring environment. XMetaL, FrameMaker+SGML, and MS SourceSafe were chosen as the authoring, publishing, and version control tools, respectively. A prototype system consisting of these tools has proven to be productive, and XML has been chosen as the standard markup language for the office. Although authors were still having difficulties adjusting to the new structured-content authoring environment, customized user screens and a new business process were designed to ease this transition. The acquisition of an electronic workflow system was put on hold while the new business process was still being reinvented and a classified local area network was being designed for security approval and implementation.

2.2 Guidance Streamlining Objective

A working group was formed to investigate how best to manage the classification guidance information, how rich the information should be, and what tools and methods to use in managing the information. The group found that classification guidance information needs to be organized by unique subject so that duplicate, inconsistent classification guidance, and gaps can be easily detected for corrective actions. Each classification topic, as a reflection of policy, needs to be linked to other information. Such information includes rationale, basis, and reasons for having the policy; how-to-apply information; related policy (or technical concepts); and keywords. This web of information defines how rich the information should be. Topic Maps have been evaluated and applied as a method for managing this information, and some of the early results are described in this paper.

§ 3 Central Publishing System Requirements

One of the early drivers for the GSI was the need for a new electronic publishing system in the ICCP. The initial requirements soon led to consideration of an XML-based system with a workflow-management system to support it. Further consideration led to the realization that managing guidance required finer granularity than just managing documents, and that realization led to examination of management at the level of the guidance topics. Discussions with internal experts and a study done by ISOGEN International led to an examination of Topic Maps as a means of managing many small, but interrelated, components.

3.1 Document Structure

A typical classification guide consists of both narrative passages and rules. The narrative portions give background for classification and establish general principles for classification. The rules, which often appear in quasi-tabular form, apply the principles to more specific circumstances and associate classifications with the rules. In some classification guides, the narrative portions may constitute several chapters, but there will be only one narrative sequence and one set of rules. In larger guides there may be many sequences of explanatory matter and related rules. The principal guide for Y-12, a typical site-wide guide, has relatively brief explanatory matter, but it has about 1,500 rules, with some brief commentary interspersed among the rules. Guides may also have components typical of technical publications, such as glossaries, lists of acronyms, and bibliographies.

Although rules may be simple statements, they often are constructed in terms of branching hierarchies. A rule may begin with the statement of a basic premise, such as the mention of a material used in weapons production. If it is generally known that the material is used somewhere in the process but not known how, the rule may branch into unclassified and classified variants for general references to the material and references in relation to a specific weapon or portion of a weapon. Frequently, the branching goes down several levels as additional conditions are specified.

The classification portion of a rule includes not only a level of classification (unclassified, confidential, secret, top secret) but also a category and potentially other information such as the conditions under which the classification can be downgraded or the reason a statement may be exempt from public release. The categories include "National Security Information," "Restricted Data", and "Formerly Restricted Data". Some rules do not lead to a single level and category but may indicate a range and direct the user of the guide to consult with a subject-matter specialist for further determination.

Although the typical guidance document can be produced with conventional desktop-publishing tools if the goal is to have only a paper document, DOE is trying to look beyond traditional publishing and delivery requirements. In particular, we do not feel it is sufficient just to present rules in the combination of outline and table form as has been done in the past. It is important that the publishing process recognize the hierarchical relationships among the rules. It is also important to select classification information only from lists of validated combinations. In the long term, we are looking also towards alternative methods of delivering classification guidance, such as the online Classification Resource System (CRS) at Y-12.

3.2 Coordination

The process of developing classification guidance tends to be interactive, involving guidance writers, classification specialists, policy specialists, and technical subject-matter experts. Guidance writers have long had a practice of collecting notes on the guidance they are writing, particularly as a reflection of their coordination with other participants in the process, both in their local operations and in the field where the guidance will be applied. The writers feel it is important to carry this coordination information along in the documentation so that future reviewers of the guidance can understand how certain decisions were made. The coordination information is not, however, generally published in the guides resulting from the process.

3.3 Versioning

The first plans for a new publishing system at ICCP included a conventional document-management system because ICCP's initial goal was simply to upgrade the publishing process. Archiving the released versions of documents, as well as tracking of revised documents through the approval process, will call for some variety of conventional workflow/document management system.

As GSI has progressed, we have realized that a much finer level of granularity is needed. Although a final design decision has not yet been made, we believe that version control should be implemented down to the level of individual guidance topics. Entire guides are revised much less frequently than there are changes to individual topics because of policy changes. While the issuing of changes is not

so common as the issuing of change pages in industrial or military documentation, it is done occasionally. A full loose-leaf publishing system is probably not required: Y-12 typically issues just a notice with the altered text. Nonetheless, it is necessary to maintain a history of changes in guidance. If, for example, a Freedom of Information Act request is made for a document created several decades ago, we need to consider what guidance was in effect at that time as well as what is current.

A topic map becomes almost a necessity for navigating the myriad objects under management if the granularity is at the level of topics. However, the techniques for managing documents at this level have been studied for a long time. An article by Han Schouten [Schouten 1989] set forth a hypertext schema for a database of document components as long ago as 1989. The details of how a hypertext version-control system will be built for this particular application is now being examined.

3.4 XML Publishing System Design

One of the earliest goals of this project, before the GSI took shape, was to provide a new publishing system for ICCP. Very early in the process, XML became an assumed requirement for any new publishing tools. While the GSI evolved and large philosophical issues related to guidance management have entered the discussion, work has nonetheless quietly continued on the publishing component.

Y-12 demonstrated several years ago that it could capture a guidance document using a fairly conventional DTD developed for technical publishing. Y-12 currently delivers several guides online in CRS over a classified network, using HTML. Neither the publishing DTD nor HTML addresses more than the presentation of the guides, however. None of the special relationships, such as rule hierarchy, is captured.

ICCP has acquired both an XML editor and an XML-aware composition system. Having developed a DTD for a general classification guide, ICCP staff have started production of several guides as a training and demonstration exercise.

Y-12 has also written a DTD as part of the analysis process. The ICCP DTD currently emphasizes the coordination and versioning data as well as providing for elements of the general publishing process, such as front matter and appendices. The Y-12 DTD, which concentrates on topical guidance as input for topic map creation, emphasizes rule structure and constraints. As the development process continues, we plan to look at DTD harmonization and/or creating scripts to convert documents between structure variants.

§ 4 GSI and a Design for Managing the Content of Classification Guides

When ICCP was early in its quest for an improved publishing technology, it realized that just upgrading the publishing tools was not sufficient. For a system to be valuable to more than just the guidance writers in Germantown, it would need to address the issues facing creators and users of guidance across DOE. DOE also wanted to reduce the number and complexity of guides in use.

To accomplish such a reduction, DOE would have to analyze the complex interrelationships among guides. Which statements in which field guides depended on which statements in master guides? Where were the best statements of classification topics? What were the essential concepts behind each topic? How could we balance redundancy of statement (in support of ease of finding a topic) with the need to maintain as few statements as possible? There are only a few very high-level reasons for classifying information (e.g., design of weapons, specialized manufacturing techniques). How could we trace the logical path from one of these "keystone" concepts to any particular piece of classification guidance?

GSI, which was begun to address some of the issues related to the interrelations among guides and to seek ways of compacting guidance, soon became intertwined with the search for a new publishing system.

One of the earliest projects in GSI was the analysis of a complex of classification topics in a very narrow and containable subject-matter area. The analysis eventually involved more than sixty

candidate topics from seven guides. After analysis, only about seven distinct topics emerged. When those of us with Topic Map experience saw the results presented in tabular form, we realized that we were seeing a paper topic map. Since Topic Map technology was already being discussed in the classification community, we decided to capture the table in XTM form and use a Topic Map browser to show how the hyperlinked representation elucidated relationships.

As a result of this experiment, we are now revising our view of the relationships that a guidance-management system will need to capture. We have now scaled up the project to explore the relationship between a large local guide and the master guides from which it is derived. In this new topic map we examined nearly 1,500 guidance topics from Y-12's principal guide and their relationships within the guide and externally to one of the authoritative guides from which the Y-12 guide is derived. From the selected source guide, we eventually mapped about 150 guidance topics.

§ 5 Topic Map for GSI

A topic map of the interrelationships among guidance documents will initially focus on the topical guidance rather than the narrative and explanatory passages because the guidance topics are discrete units. Documentation should exist for local and derived guides to validate each guidance topic by reference to a corresponding topic in a master guide or some other authoritative source.

We currently believe, however, that direct mapping of a topic in one guide to a topic in another guide will not be sufficient to enable all the goals of GSI. For the current paper documents, it is sufficient to provide documentation of authority. But the essence of making classification decisions does not depend only on specific statements but rather on the concepts behind those statements. Accordingly, it looks more fruitful to map interrelationships in terms of concepts. The original GSI studies did attempt to isolate the key concepts behind classification topics, calling them "keywords" and "keystones." Keywords may be actual terms used in the guidance topics, or they may be slightly abstracted terms that capture the essence of the topics. In effect, they are like the trigger concepts at the top of the implication networks for Y-12's classification engine, Ferret [Mason 2002]. Keystones are more abstract than keywords; they represent the high-level concepts that are the justification for classifying information. There will likely be many keywords but only a few keystones. It should be possible to create Ferret-style implication networks leading up from keywords to keystones.

In the early GSI study, the initial list included about sixty primary topics. Although keyword analysis was not done on the entire list, Y-12 estimates over a hundred keywords would be needed. The original XTM demonstration examined in detail seven classification topics, with about thirty instances. The seven topics reflected combinations of seven keywords but only two keystones. However, the early study — and the topic map that reflected it — focussed on direct relationships among guidance topics, not patterns of keyword use, which might not emerge from so small a sample.

The new topic map for the Y-12 classification guide and one of its sources has nearly 3,000 keywords with its 1,650 guidance topics. Not all guidance topics (e.g., topics whose result is unclassified) have been assigned keywords. Because the selection of keywords has not been fully harmonized between the guides, there may be some reduction of the final number. This collection of keywords also includes much of the Ferret implication trees, so not all of the collection is involved in the linkages between the two guidance documents currently in the topic map.

We expect that when keyword analysis is done on larger collections of guidance topics, it will be possible to map the dependencies among topics automatically according to their patterns of keyword use. If that is achieved, we will have found a means for consolidating topics and thus of streamlining the amount of guidance that must be managed.

5.1 Ontology for Guidance Management

Because of its specialized field of application, a full topic map for GSI must deal with a unique ontology of objects and relationships among them.

5.1.1 Objects

Just to enable a topic map to operate on guidance data, many classes of topics must be created and populated. Some of these classes exist to type other topics, such as guidance documents and the guidance topics within them. Typing topics are relatively few in number. Other topics serve as indicators of static properties, such as classification values. These, too, are relatively few in number. A third relatively small class of topics will serve to type roles in associations.

Much more numerous will be the topics used as proxies for the classification guides and the topical guidance within them. Along with these we expect one or more classes of topics for coordination information and a similar group of classes for versioning, tracking, and managing topics.

5.1.1.1 Guides

The ADC traditionally begins classification analysis of a candidate document by applying a classification guide. In the original GSI study and topic map, no distinction was made between master guides and local guides. However, for GSI to achieve its goals, it will be necessary to make such distinctions. "classification guide" is nonetheless a useful concept as a superclass within the overall ontology, with subclasses "master guide" and "derivative guide." The distinction between master and derived guide is not absolute. For example, one of the guides frequently cited as an authority in the principal local guide and Y-12's other local guidance sets official policy for a wide range of weapons-related issues. But it is itself derived in many areas from other high-level guides which set policy at an even level near what could be considered keystone. At present there are only two levels of guide, master and derived, in the topic map because within a single association that appears to be sufficient. In the future, as the number of layers of derivation is increased, additional typing topics that reflect degrees of authoritativeness/derivation may be created.

5.1.1.2 Guidance Topics

The usual focus for applying guidance is the guidance topic (not to be confused with a topic-map topic!). In a topic map for managing guidance, "guidance topic" will be another superclass. Within the class are master and derived topics. The ontology also provides for hierarchies of topics, so typing topics for roles such as parent and child have been created. (Topics for typing siblings may emerge, but in practice, sibling topics can be derived rather than explicitly notated.)

An instance for an actual topic is then established by reference to one of these typing topics. For convenience in browsers that highlight basenames, the text of the guidance topic is stored in the baseNameString (here obfuscated because most real guidance topics are themselves classified). This topic is typical in that there is a principal statement that branches into two qualifications on the original statement.

```
<topic id="gtd-BakingGuide-1-1">
  <instanceOf>
    <topicRef xlink:href="#r-topic-in-local-guide" />
  </instanceOf>
  <baseName>
    <baseNameString>Use of both baking powder and baking soda in a
single recipe</baseNameString>
  </baseName>
</topic>
```

```
<topic id="gtd-BakingGuide-1-1.1">
  <instanceOf>
    <topicRef xlink:href="#r-topic-in-local-guide" />
  </instanceOf>
  <baseName>
    <baseNameString>Product unspecified</baseNameString>
  </baseName>
</topic>
```

```

<topic id="gtd-BakingGuide-1-1.2">
  <instanceOf>
    <topicRef xlink:href="#r-topic-in-local-guide" />
  </instanceOf>
  <baseName>
    <baseNameString>Product specified</baseNameString>
  </baseName>
</topic>

```

(A more complete version of an instance is shown below.)

5.1.1.3 Keywords

As keys to identifying the significant content of classification topics, keywords will be major components of topic maps. A high-level typing topic establishes a superclass of concepts, with subclass typing topics of keystones, keywords, and components in Ferret-like implication networks: antecedents and consequents for logical associations.

It may be useful to examine guidance topics according to their whole list of keywords or trigger concepts; a typing topic for these also appears in the ontology. At this stage of the analysis of the actual subject matter, we are not sufficiently advanced to decide whether it is useful to create a new class of concept groups.

Because keyword hierarchies form implication networks, topics to identify roles in implication, such as antecedent and consequent, have been defined. The top of the implication hierarchy may well turn out to be the collection of keystones, or it may be elements from the Master Subject List. At this time the implication hierarchy is still under development; however, the pattern will probably turn out to be like that used in the Ferret system [Mason 2002]. Individual implications are built up of antecedents and consequents. The antecedent and consequent roles are, of course, relative to individual associations in an implication tree that is several layers deep (some Ferret trees are as much as a dozen levels deep). A high-level participant in an implication tree may turn out to be equivalent to a keyword associated with a guidance topic. Ferret trees in previous applications work up from the level of terms used in actual documents to the level of trigger concepts for rules. These triggers seem to be equivalent to keywords. The keywords, in turn, are the bottom level of trees reaching up towards the keystones. The resulting implication network will span the entire range from high-level policy abstractions down to the concrete language used in actual documents.

5.1.1.4 Keystones

As keys to identifying the reasons for classification of subjects at a high level, keystones will also need a typing topic. Keystones are at such a high level in the logical hierarchy that they are not likely to need further typing topics. Although a few keystones were included in the earliest GSI topic map, none have yet been identified for the larger example that links two guides. Some keystones may emerge from analysis of the guide that sets top-level policy for classification. A "Master Subject List" that has been under development by another part of the GSI project may also provide keystones, as well as parts of the trees connecting them with the keywords for guidance topics.

5.1.1.5 Coordination and Reference Items

Because one of the major purposes of building a topic map for guidance management will be to help authors and reviewers of guidance documents, the topic map will need to provide locations for coordination information. Such information constitutes a form of metadata about guidance topics; the details of its form are not yet evident. The DTD being developed in conjunction with the publishing system is only beginning to define these elements; the ontology in the topic map will probably follow what develops in the publishing system. Because these items may contain extensive text, we shall have to find a means for presenting structured text in a topic-map browser.

5.1.1.6 Properties Associated with Guidance Topics

Topics are also generally numbered and are frequently referenced by number. Although numbering is not an intrinsic part of a topic, it appears in the topic map for identification reference purposes as

a typed and scoped name in addition to the primary name applied to each guidance topic. In a topic map, hyperlinks and names are the primary means of locating information. Because topic numbers sometimes change when guides are revised, versioning, probably through scopes, will be applied to topic numbers as well as to the texts.

Guidance topics usually consist of two significant parts: the statement of a condition and the classification information associated with that condition. So that they may be referenced elsewhere in the map, typing topics for the components of classification are provided, and then instances of these typing topics appear as static topics that represent the allowable values or ranges of values of classification items. There are values for the level of classification, the category of classification applied, and, for use with national security information, further definition of the types of exemption for release and circumstances under which the information can be declassified. One section of the current topic map is devoted to associations between individual guidance topics and their respective classification components. Because some guidance topics may cover more than one kind of classification, the association provides for both primary and secondary level and category.

5.1.1.7 Source Documents

Because the topic map is intended to manage a collection of documents and components, it will have many pointers to occurrences in external storage. The storage system is just beginning to evolve, but we have provided topics to scope occurrences, such as XML sources and HTML and PDF renditions. Source references will be further qualified by the versioning system. The occurrences will likely be of both the individual items under configuration control (e.g., guidance topics) and the resulting completed documents that accumulate the items.

5.1.2 Association Roles and Links

Associations are what make a topic map different from most other collections of metadata. In the GSI topic map we capture a web of relationships that as a whole would be difficult to capture or manipulate in more traditional metadata systems. Guidance topics appear in logical hierarchies within their parent guides. Derived guidance topics must also be validated by association with topics in master guides, and these topics have their own hierarchy. Keystones and keywords will occupy their own hierarchies of topics and associations. Because we intend to use keywords and keystones to manage guidance, associations between guidance topics and these conceptual topics will form the core of the actual map.

5.1.2.1 Links around Guidance Topics

Compilers of classification guidance have traditionally thought of direct links among guidance topics as being the primary connection between corresponding portions of guides. Validation of a guidance topic in a derived guide has always been by association with a guidance topic in a master guide. For purposes of official approval, such direct links will probably continue to have major significance.

Direct links are grounded only in an assertion by a guidance writer. In most cases, human inspection will validate the content of a link. However, such validation cannot be confirmed by a computer. For the purposes of managing guidance through a topic map, other links, such as those based on keywords, will be necessary.

A topic rarely stands by itself within a single guidance document; it is usually part of a hierarchy within the document. In some guides, such as Y-12's main guide, the hierarchy is relatively flat. Much of the guide is devoted to materials used in manufacturing, and the hierarchy rarely goes deeper than rules about specific materials, such as steel or hydrogen. Some rules in the Y-12 guide branch, as they do in many other guides, depending on whether a specific product is mentioned or not. Other guides, such as the DOE-wide guide on safeguards and security, have deep outline structures that go down more than six levels.

The immediate environment of a given rule can be seen as a complex of potential links not unlike those in a thesaurus: broader topic, narrower topic, related topic. Display in a topic map browser does not require making all these relationships explicit: simple parent-offspring links are sufficient

to assemble most other relationships. We are only beginning to develop and exploit such a wider set of relationships.

The root of a hierarchy of guidance topics is generally a guidance document. Because this system is intended to maintain guides, this hierarchy becomes particularly important. For the set of sample topics above, the hierarchy could be represented by the association:

```
<association>
  <instanceOf>
    <topicRef xlink:href="#f-topic-to-parent" />
  </instanceOf>
  <member>
    <roleSpec>
      <topicRef xlink:href="#r-classification-topic-parent" />
    </roleSpec>
    <topicRef xlink:href="#gtd-BakingGuide-1-1" />
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#r-classification-topic-child" />
    </roleSpec>
    <topicRef xlink:href="#gtd-BakingGuide-1-1.1" />
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#r-classification-topic-child" />
    </roleSpec>
    <topicRef xlink:href="#gtd-BakingGuide-1-1.2" />
  </member>
</association>
```

A topic in a derived guide may, first of all, be established by an identification topic, as indicated above. Then its relationship to a master topic, established similarly, can be shown in an association. So, if the topic about baking powder, in a derived guide, is based on one about leavening, in a master guide, the relationship might be shown as follows:

```
<association>
  <instanceOf>
    <topicRef xlink:href="#f-mainTopic-to-instance-topic" />
  </instanceOf>
  <member>
    <roleSpec>
      <topicRef xlink:href="#r-topic-in-local-guide" />
    </roleSpec>
    <topicRef xlink:href="#gtd-BakingGuide-1-1" />
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#r-topic-in-master-guide" />
    </roleSpec>
    <topicRef xlink:href="#gtm-Leavening-315" />
  </member>
</association>
```

Here, the derived topic (gtd-BakingGuide-1-1) is associated with a master topic (gtm-Leavening-315). Each guidance topic has been defined elsewhere in the topic map.

As rules for determining classification, most guidance topics have an associated classification unit, which can also be attached to a topic with an association.

```
<association>
  <instanceOf>
    <topicRef xlink:href="#f-topic-classification" />
  </instanceOf>
  <member>
```

```

    <roleSpec>
      <topicRef xlink:href="#r-classification-topic" />
    </roleSpec>
    <topicRef xlink:href="#gtd-BakingGuide-1-1.2" />
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#r-classification-category" />
    </roleSpec>
    <topicRef xlink:href="#cat-RD" />
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#r-classification-level" />
    </roleSpec>
    <topicRef xlink:href="#lev-C" />
  </member>
</association>

```

Here it is established that the topic on using both baking powder and baking soda in the recipe for a single product (#gtd-BakingGuide-1-1.2, defined earlier) has associated with it the classification level and category of "C-RD". (While the association of both baking soda and baking powder with a given recipe — say, buttermilk biscuit — as an example of confidential restricted data sounds improbable on the surface, it corresponds to what real guidance topics might indicate concerning the association of a particular material with a real weapon.)

Most of the association instances having to do with interrelationships among guidance topics and guidance documents can be populated from existing documents. The demonstration topic map was created by simply running scripts against XML sources.

5.1.2.2 Links Involving Topics with Keystones and Keywords

Computer-aided management of guidance needs more than just the assertion that guidance topics are similar: it needs means for representing such similarity. The most promising approach that is readily available is to associate keywords (concepts) with guidance topics, then look for patterns of keywords. If two topics are both represented by the same group of keywords, they must refer to the same subject matter, whether their text appears to be similar or not.

Keystones provide a conceptual framework above both keywords and guidance topics. If the typing topics in the topic map provide an ontology of objects related to the classification process, the keywords should provide the basis for a taxonomy. The GSI team and others in the DOE classification community have been working on a "Master Subject List" that can provide a starting point for such a taxonomy. When fully developed, such a taxonomy should provide a logical path from the keystones down to the individual guidance topics. In building such a full taxonomical hierarchy, techniques similar to those used in the implication networks of the Ferret system could be used [Mason 2002].

The basic pattern of association is as shown below, where the guidance topic "gtd-BakingGuide-1-1" is associated with three keywords, representing the recipe and the two leavening agents. The two branching topics would then have subsidiary associations that link the parent topic with individual qualifiers. This association would allow direct access to guidance topics through keywords, taken individually. A simple browser interface can lead from a single keyword to the other keywords that appear in associations with it and thence to guidance topics. The primary association involves only the guidance topic and the three keywords.

```

<association>
  <instanceOf>
    <topicRef xlink:href="#f-mainTopic-to-keyword" />
  </instanceOf>
  <member>
    <roleSpec>
      <topicRef xlink:href="#r-topic-in-local-guide" />
    </roleSpec>

```

```

    <topicRef xlink:href="#gtd-BakingGuide-1-1" />
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#r-keyword" />
    </roleSpec>
    <topicRef xlink:href="#kw-recipe" />
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#r-keyword" />
    </roleSpec>
    <topicRef xlink:href="#kw-baking_powder" />
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#r-keyword" />
    </roleSpec>
    <topicRef xlink:href="#kw-baking_soda" />
  </member>
</association>

```

The two branching associations connect the parent topic with the appropriate keyword.

```

<association>
  <instanceOf>
    <topicRef xlink:href="#f-mainTopic-to-keyword" />
  </instanceOf>
  <member>
    <roleSpec>
      <topicRef xlink:href="#r-topic-in-local-guide" />
    </roleSpec>
    <topicRef xlink:href="#gtd-BakingGuide-1-1.1" />
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#r-classification-topic-parent" />
    </roleSpec>
    <topicRef xlink:href="#gtd-BakingGuide-1-1" />
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#r-keyword" />
    </roleSpec>
    <topicRef xlink:href="#product_unspecified" />
  </member>
</association>

```

```

<association>
  <instanceOf>
    <topicRef xlink:href="#f-mainTopic-to-keyword" />
  </instanceOf>
  <member>
    <roleSpec>
      <topicRef xlink:href="#r-topic-in-local-guide" />
    </roleSpec>
    <topicRef xlink:href="#gtd-BakingGuide-1-1.2" />
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#r-classification-topic-parent" />
    </roleSpec>
    <topicRef xlink:href="#gtd-BakingGuide-1-1" />
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#r-keyword" />
    </roleSpec>
  </member>
</association>

```

```
</roleSpec>
  <topicRef xlink:href="#product_specified" />
</member>
</association>
```

Such an association establishes only the logical antecedents or trigger concepts for a single guidance topic. This mechanism could be extended, however, in the process of managing guidance, particularly if the goal is to streamline by reducing redundant topics. The first step is to create associations that just combine sets of keywords and provide each associated set with an identifier. A guidance topic would then be associated not directly with keywords but rather with one of the established groups. Examination of the topic map for all topics playing the role of "r-classification-topic" in association with a single keyword group would then show potential redundancies.

Finding guidance topics according to their keywords is only one part of managing guidance. A major task for guidance creators is finding all the topics that must change when there is a change in policy or the environment under which classification is employed. Classification is always contextual; it depends partially on the details of information under consideration, partially on the immediate context of the information, and partially on the overall context of what is known at the unclassified level. If any one of those changes, the classification may change. If, for example, a treaty changes what is known at the unclassified level, then guidance managers must work down from the change in unclassified knowledge to the topics.

The path from a high-level concept, such as a keystone or a relatively abstract keyword, to the keywords associated with guidance topics will form part of an implication network. The use of such networks in a classification knowledge base has already been demonstrated in [Mason 2002]. The knowledge base is built up of many associations that bring together keywords as antecedents and consequents.

```
<association id="implication-05">
  <instanceOf>
    <topicRef xlink:href="#f-rule" />
  </instanceOf>
  <scope>
    <topicRef xlink:href="#layer-yeastless_bread" />
  </scope>
  <member>
    <roleSpec>
      <topicRef xlink:href="#role-consequent" />
    </roleSpec>
    <topicRef xlink:href="#ks-leavening" />
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#role-antecedent" />
    </roleSpec>
    <topicRef xlink:href="#kw-baking_powder" />
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#role-antecedent" />
    </roleSpec>
    <topicRef xlink:href="#kw-baking_soda" />
  </member>
</association>
```

This implication association suggests that keystone "ks-leavening" sits above two different paths up through the network, those leading through keywords "kw-baking_powder" and "kw-baking_soda". The scope, "layer-yeastless_bread" is introduced into this association because we have discovered that the taxonomy of the Master Subject List and similar analyses of our subject-matter area cannot

be simple trees. That is to say, for example, that "kw-baking_soda" implies "ks-leavening" in the context represented by "layer-yeastless_bread" and perhaps not in others.

While such an implication network is similar in structure to that used by the Ferret system, it works at a different level. This network traces the path from very abstract concepts, the keystones, down to a much less abstract level, the groups of keywords that represent the essence of individual guidance topics. The Ferret network begins at that level of trigger concepts for guidance topics and works down to the very concrete level of language that can be recognized in documents that are being analyzed for classification. The similarity of the networks means, however, that the same tools — such as browsers and editors — can be used on both. Furthermore, the two networks can be merged to provide a seamless path of traceability from policy to its application to individual documents.

5.2 Instance Pointers

5.2.1 Guides

The actual reference copies of the classification guides under management will be stored as external resource files. Proxy topics for the guides must be created within the map, with pointers to the external resources. A topic establishes the reference point for an actual guide, creates an ID for future references, and establishes the identity of the guide through reference to an XML source file for the assembled document. Guides are frequently referenced only by a report number because their formal titles tend to be long and cumbersome; the baseName is structured accordingly. Besides the source that may maintain the official master text of a guide, there may be other manifestations, such as a display version in HTML, as indicated by a scoped occurrence.

```
<topic id="gtd-BakingGuide-1">
  <instanceOf>
    <topicRef xlink:href="#r-guide" />
  </instanceOf>
  <subjectIdentity>
    <subjectIndicatorRef xlink:href="BakingGuide-1.xml" />
  </subjectIdentity>
  <baseName>
    <baseNameString>BakingGuide-1</baseNameString>
    <variant>
      <parameters>
        <subjectIndicatorRef xlink:href="#r-full_name" />
      </parameters>
      <variantName>
        <resourceData>Guide for Baking Systems</resourceData>
      </variantName>
    </variant>
  </baseName>
  <occurrence>
    <instanceOf>
      <topicRef xlink:href="#r-source" />
    </instanceOf>
    <scope>
      <topicRef xlink:href="#s-html" />
    </scope>
    <resourceRef xlink:href="BakingGuide-1.html" />
  </occurrence>
</topic>
```

5.2.2 Instance Guidance Topics

Every guidance topic that is under management must be represented by a topic-map topic. Each such topic will form a hub around which data is assembled, either as actual data in name strings or through associations with guides, keywords and keystones, classification values, and other related topics. If master copies are kept as resources outside the map, pointers must be created to the resources.

An individual topic in a guide has a baseName in the scope of its topic number (subject, possibly to a versioning scope) and another baseName that is a proxy for its text (as shown above). The topic also has a pointer to a location in a file containing the data source for the guidance topic:

```
<topic id="gtd-BakingGuide-1-1">
  <instanceOf>
    <topicRef xlink:href="#r-topic-in-local-guide" />
  </instanceOf>
  <baseName>
    <scope>
      <topicRef xlink:href="#gd-BakingGuide-1" />
      <topicRef xlink:href="#r-fulltext" />
    </scope>
    <baseNameString>Use of both baking powder and baking soda in a
single recipe</baseNameString>
  </baseName>
  <baseName>
    <scope>
      <topicRef xlink:href="#gd-BakingGuide-1" />
      <topicRef xlink:href="#r-topicnumber" />
    </scope>
    <baseNameString>1
  </baseNameString>
</baseName>
  <occurrence>
    <instanceOf>
      <topicRef xlink:href="#r-source" />
    </instanceOf>
    <scope>
      <topicRef xlink:href="#s-xml"/>
      <topicRef xlink:href="#gd-BakingGuide-1"/>
    </scope>
    <resourceRef xlink:href="BakingGuide-1.xml#I1"/>
  </occurrence></topic>
```

The creators and users of classification guidance are accustomed to seeing text. The "baseName" mechanism in XTM was included to provide representations of topics that would be visible to the users of Topic Map browsing tools. Most tools, accordingly, provide featured presentation of names. The creators of XTM may have thought of base names as labels for topics, but we are using name strings as visible proxies for repositories for text. Thus our names tend to have whole sentences; once browsers support namespace-qualified extensions to the XTM DTD, even tagged text may show up in baseName strings. The result is that our topic maps are not merely collections of metadata: they display real data as well.

§ 6 Using the Map

The demonstration topic map is only beginning to accumulate enough links to support managing guidance. Even in its preliminary state, however, we are learning new things about existing guidance. Although the dependency information in the map comes from existing sources, the ability to bring the related topics together in a browser is allowing us to consider dependencies more thoroughly. The process of supplying keywords, like the process of building the Ferret knowledge base before it, is causing us to rethink some of the derived guidance topics.

The topic map for GSI is still an experimental construct that will evolve as the overall system design evolves. However, we are already working on an overall architecture in which the topic map will represent the primary management structure.

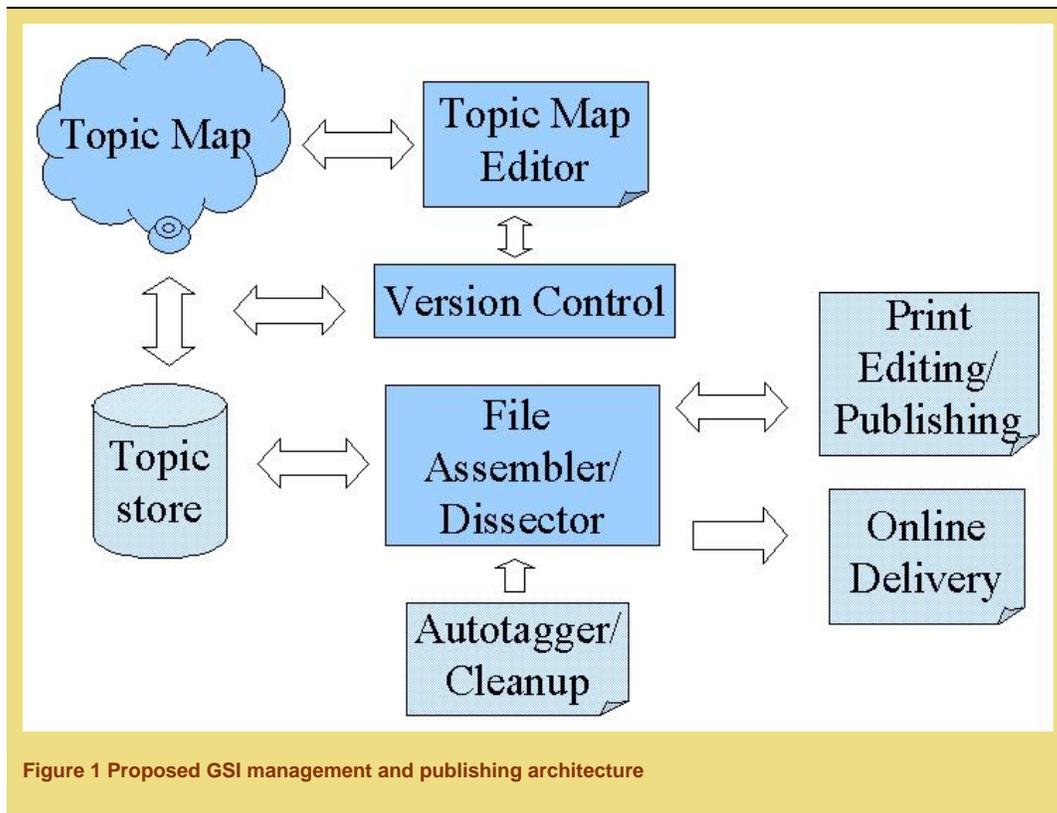


Figure 1 Proposed GSI management and publishing architecture

This architecture provides for a dual user interface, through a topic-map browser/editor and through the publications system already being prototyped. The interface through the topic map will give access to the full range of data and functions in the system, including both text editing and the versioning and document-management system. The publications interface builds on the strengths of commercial publications products, not only to provide printable output but also for such well integrated functions as spelling checking. Between the two sides of the system is a component that assembles complete documents for the publishing system out of the stored fragments or dissects whole documents that have been changed in the publishing system into storable fragments. We are also looking for software to assist the capture of the hundreds of legacy guides across DOE.

§ 7 Future Directions for Guidance Management

Having demonstrated that it is possible to scale up this sort of topic map to thousands of guidance topics and keywords, in many hundreds of relationships (the resulting XTM file for the demonstration is over 200,000 lines long), we are looking forward to both extending it and to putting it in the hands of potential users.

Y-12 is thinking of adding more master guides to the map. Y-12 classification writers already know the basic relationships because they were documented in the process of validating and getting approval for our local guides. Because we have scripted the generation of some portions of the map, the most time-consuming part of adding a new guide is assigning keywords to the guidance topics.

With the proof-of-principle part of the project behind us, we are now looking at improving the user interface to the system. We have heretofore used a commercial topic-map browser whose generic interface was not designed to highlight the particular components and relationships in this map. We have been able to demonstrate the map, and we already have feedback from the user community about things they would like to see presented differently. We are now designing new HTML and JSP pages to replace the generic interface. As part of the Ferret project, Y-12 developed a simple

editor for knowledge bases. Because of the similarity between the Ferret knowledge base structure and this topic map, we are modifying the editor to write out topic-map components. The next step will probably be to implement a topic-map editor, since the files in question are not conducive to easy editing in a conventional XML editor.

In the next year, besides the considerable redesign of the interface, we expect to start on the implementation of the content-management system and the extensions to both the topic map and the interface to support it. We already have requests from sites in the field for access to the topic-map system. If both it and the publishing system prove successful, we may have the means for spreading integrated guidance management and publishing across DOE.

Acknowledgements

The authors of this paper would like to acknowledge the contributions of the DOE working groups and consultants whose knowledge and experience have set the course of the Classification Guidance Streamlining Initiative. In particular, the continued support and attention this initiative receives from the management of DOE offices of Information Classification and Control Policy and Security Policy Staff were extremely valuable, in times of scarce resources and competing projects. Many thanks are due to the director of Technical Guidance division, Dr. Andrew P. Weston-Dawkes, who has been involved in making every key decision.

The analysis performed by ISOGEN International (at the time affiliated with DataChannel) and particularly by Dr. Steven R. Newcomb has been very influential in shaping the overall design of the system.

The initial prototype publishing system developed by Todd Powell and Rebecca Dahlman, of SOZA, Inc., was instrumental in determining the business process.

Dr. Peter J. Kortman, of the Y-12 Classification Office, has been the driving force behind the development of the Ferret system, created by Robert McGaffey and Michael Bell. Work on the Ferret knowledge base led to the initial application of Topic Maps to classification guidance. Richard Baylor, head of the Y-12 Classification Office, has provided support and leadership since the earliest days of the CRS and Ferret projects.

The Y-12 National Security Complex is managed for the U.S. Department of Energy by BWXT Y-12, L.L.C., under contract DE-AC05-00OR22800.

This document was prepared as an account of work sponsored by an agency of the U.S. Government. Neither the United States Government nor any agency thereof, nor Contractor, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, use made, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency or Contractor thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency or Contractor thereof. Further, BWXT Y-12 is not responsible for the contents of any off-site pages referenced.

This document was prepared by a contractor of the U.S. Government under contract DE-AC05-00OR22800. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce these documents, or to allow others to do so, for U.S. Government purposes. These documents may be freely distributed and used for non-commercial, scientific and educational purposes.

Bibliography

- [**DOE 2001**] Department of Energy, Information Classification and Control Policy, *User Requirements for the Classification Guidance Database and Publishing System* (unnumbered specification), March 19, 2001
- [**ISOGEN 2001**] ISOGEN International/DataChannel, Inc., *High-level System Design Overview* (response to specification), April 24, 2001
- [**Mason 2002**] Mason, James David. "Ferrets and Topic Maps." *Markup Languages: Theory and Practice* 3 No. 2 (Spring 2001): 123-140 .
- [**Schouten 1989**] Schouten, Han. "SGML*CASE: The Storage of Documents in Databases." *SGML Users' Group Bulletin* 4 No. 1 (1989): 1-14.

The Authors

Vinh Lê

U. S. Department of Energy, Office of Security, Information Classification and Control Policy (SO-12)
Germantown
MD

Vincent Dinh Vinh Lê was formally trained as an electrical engineer and a systems engineer. During the late 1980s, he was a systems operations engineer working for the Potomac Electric Power Company (PEPCO). He modeled PEPCO's electric systems and wrote operating procedures for the safe, reliable, and economical operations of the electric systems. He managed a database which was used by PEPCO in "rolling blackouts". In 1991, he joined the Department of Energy (DOE) where he was trained as a nuclear safety engineer and led teams assessing the implementation and effectiveness of a DOE-wide incident reporting system. Mr. Lê^a has developed and written information classification guidance pertaining to nuclear weapons and materials disposition activities, nuclear arms control and treaty issues, and DOE safeguards and security. Currently, he leads an initiative to develop a modern DOE classification guidance system applying XML and Topic Maps standards.

James David Mason
Y-12 National Security Complex
Bldg. 9113, M.S. 8208
Oak Ridge
TN
37831

James D. Mason, originally trained as a mediaevalist and linguist, has been a writer, systems developer, and manufacturing engineer at U.S. Department of Energy facilities in Oak Ridge since the late 1970s. In 1981, he joined the ISO's work on standards for document management and interchange. He has chaired ISO/IEC JTC1/SC34, which is responsible for SGML, DSSSL, Topic Maps, and related standards, since 1985. Dr. Mason has been a frequent writer and speaker on standards and their applications. For his work on SGML, Dr. Mason has received the Gutenberg Award from Printing Industries of America and the Tekkie Award from GCA. Dr. Mason was Chairman of the Knowledge Technologies 2002 conference sponsored by IDEAlliance. He is currently working on information systems to support the classification community at DOE's Y-12 National Security Complex (Y-12) in Oak Ridge, Tennessee.

Extreme Markup Languages 2002

Montréal, Québec, August 6-9, 2002

This paper was formatted from XML source via XSL

Mulberry Technologies, Inc., August 2002