

## **Fine-grained publications management under Topic Map control**

---

Vinh Lê  
*U. S. Department of Energy*

James David Mason  
*Y-12 National Security Complex*

---

### **Abstract**

The DOE [Department of Energy] 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. As we assemble tools to manipulate both the documents and the additional metadata contained in topic maps, we are creating an unusual suite for both document and content management.

# Fine-grained publications management under Topic Map control

## *Table of Contents*

Publishing in transition.....	1
The origins of the Guidance Streamlining Initiative.....	1
Metadata and annotation.....	1
Approaches to documents and content.....	2
Topic Maps for GSI.....	3
Architecture for document management.....	4
Dissector/Assembler.....	4
Metadata editing.....	7
Content management.....	8
Future directions.....	9
Acknowledgements.....	9
Bibliography.....	10
The Authors.....	11

# Fine-grained publications management under Topic Map control

Vinh Lê and James David Mason

## § Publishing in transition

Since its origins in the Manhattan Project, DOE has had to deal with classified information. At the front line of protection of such information are the ADCs [Authorized Derivative Classifiers], who provide the initial review of information close to its sources. An ADC, whose authority is derived from those who set classification policy, depends on approved, published guidance documents to make classification decisions. With more than half a century of publishing classification guidance, DOE has a corpus of many hundreds of documents both at headquarters and at its many field locations, research laboratories, and production facilities. As we have explained in past documents [Mason 2002], managing this body of information is more complex than simply maintaining documents under workflow and version control. There are complex dependencies among the documents, so that a change in policy may have a ripple effect not only across documents but also across sites and organizations.

Classification guidance documents are in some ways typical technical publications, with nested sections and apparatus like lists of abbreviations and references. However, at a fine level of granularity they have a unique structure related to “topical guidance” that is simultaneously tabular and outline-like in appearance. The outline form is a reflection of a hierarchy, sometimes more than half a dozen layers deep, of “guidance topics.” The tabular appearance of the outline items reflects the common structure of the topics: a topic number, a statement, and, at least for the terminal nodes of the hierarchy, an associated classification value. Because of the structured nature of the documents, the ICCP [Information Classification Control Policy] organization in DOE decided several years ago to assemble an XML-based publication system to replace the word-processing software that had been used for some years.

If these documents were self-contained, the mechanics of XML tagging and publication could be handled with fairly conventional publishing tools. Both ICCP and Y-12 had independently developed XML DTDs for classification guides, and both had used COTS XML tools to produce guides and derive several types of products from the tagged guides.

### ***The origins of the Guidance Streamlining Initiative***

Very early in the analysis leading up to the creation of a system, ICCP realized that it was not sufficient to do publishing of isolated documents. There are too many dependencies among guidance documents for any one to be considered in isolation. Furthermore, converting to a new publishing system offered an opportunity to rethink documents. 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].

One of the consequences of this rethinking was the beginning of the GSI [Guidance Streamlining Initiative]. An early project to trace dependencies among a constrained set of topics spread across several guides led to an exercise in creating a paper “topic map.” This paper study led to its conversion to a real topic map, and then a later study built a topic map involving all the dependencies between a Y-12 guide and one of the master guides from which it was derived. The results of some of the new work was presented at Extreme Markup Languages 2002 [Lê and Mason 2002].

### ***Metadata and annotation***

As ICCP gains experience both with XML publishing tools and with GSI and topic maps, the development team continues to consider what functionality an overall system needs to provide. Most of the new additions concern information that is related to the creation and maintenance of classification guidance even though it is not part of any guide document.

The original GSI topic maps already contained some of this information, notably *keywords* and *keystones*. A keystone for a guidance topic is the concept that the topic is attempting to protect. There are relatively few of these, just as there is a relatively small set of reasons behind classification policy. Keywords are the distilled essence of a classification topic. In the Ferret classification application, keywords are the trigger concepts that fire individual rules. In GSI, keywords may be the means of determining dependencies, overlaps, and redundancies among classification topics. In the long run, keywords should become a major means for defining the “roadmaps” for managing and preserving knowledge for information classification policy.

The derivation, or *basis*, for guidance has always been a major concern of guidance management, and it, too, was part of the GSI topic maps, from the earliest paper studies on. All guidance must be traceable to primary policy, and it has been normal practice to document derivation in tables that trace the authority of each guidance topic back to a statement in a higher-level document. As the knowledge base surrounding guidance increases, it may be possible to supplement this data with derivation through complexes of keywords or other metadata.

Guidance authoring does not exist in a vacuum. Not only must guidance writers consider policy, they must reflect on actual practice in the field. Behind any guidance manual there is usually a body of author’s notes that reflect discussions with both classification practitioners in the field and subject-matter experts at the design and manufacturing sites associated with DOE’s programs. Some of this collaboration information may deal with large matters of policy that cover whole sections of guides, while other information may be confined to the interpretation of a single guidance topic. If collaboration data was applicable only at specific locations, it might be possible to encapsulate it into documents at those locations. However, direct incorporation does not seem appropriate for data that may apply to the whole of large units. Accordingly, we believe that it is better to handle collaboration information in an extratextual manner, through scoped associations in a topic map.

A second type of meta-information has been called *genealogy*: this is not simply an application of versioning, as might be generated by a conventional document-management system. It is concerned with conceptual content as much as it is with text strings that might — or might not — be susceptible to string comparison. Derivation, which is already considered to be in the topic map, also contributes to this information. In a sense, what genealogy deals with is not just history but subject identity, in the topic maps sense. If two currently active guidance topics have the same genealogy, they may be considered occurrences of the same topic, no matter what their current wording.

Besides keystones, a guidance topic may also have a *rationale* to justify its associated classification. We have recognized that there are two different kinds of rationales. In the case of guidance topics that point to unclassified data, there is a fairly constrained list of reasons, such as that the information is widely known or that the subject has been formally declassified. For topics that identify classified information, however, there may be more extensive explanation, including references to keystones. Either of these cases might be handled directly in a document, but they seem to call for different techniques (selection from a fixed list of attribute values vs. a piece of variable element content). A topic map provides a way around this difficulty because associations can be built between a guidance topic and one of the fixed rationales in the shared ontology, a separate topic that represents an explanation, or, in some special cases, both.

Other metadata may include *background* and *related information*. We already recognize that this metadata will probably consist of hyperlinks between the content of guidance documents and outside reference materials, ranging from the minutes of review panels to a comprehensive encyclopedia and thesaurus of weapons information being generated by DOE’s Office of Scientific and Technical Information. Such information clearly does not belong in any one guidance document; a topic map provides the best way of linking it to guidance.

*Usage* information likewise is a good candidate for linking through a topic map. Often this information will occur in separate documents, such as collections of classification scenarios used in training new ADCs. It may even be in forms that cannot be incorporated into a guidance document, such as video clips of training or demonstrations.

### **Approaches to documents and content**

Perhaps the greatest difficulty posed by the combination of requirements — document management for publication and content management for GSI — is the need to deal with different views of the information. The guidance authors prefer to see the text they are working on in context, as guidance

topics surrounded by other topics, whether those other topics are parents, children, or siblings of the ones under consideration. ICCP has already stipulated a requirement for a document-centric approach. Maintaining an audit trail for derivation and maintenance of topics, however, involves pieces of multiple documents. Annotation and metadata likewise reach outside the individual guidance document.

Our challenge is thus to maintain two views of guidance information, so that it may be approached either through whole documents (or large chunks of documents) or through the network of information surrounding classification guidance. Viewed from a Topic-Maps perspective, this is not an overwhelming problem. The network of information is all about links, and topic maps excel at collecting links. However, assembling the software to support the dual view of guidance is more complex than just acquiring a topic-map engine/browser.

We believe that in addition to the already existing publications tools and topic-map browser, we need a metadata editor, a software hub that mediates between document-centric and topic-centric approaches to information, and a place to store information in both document and topic forms.

## § Topic Maps for GSI

At Extreme Markup Languages 2002, we presented the general design for a topic map for GSI. The essential design of its ontology has not changed, but experience with building instances and with designing metadata has caused us to expand that ontology.

The ontology to enable a topic map to operate on guidance data requires that many classes of topics 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 for this primary content 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 serves to type roles in topic-map associations.

Much more numerous will be the topics used as proxies for the classification guides and the topical guidance within them. To these we add several classes of topics to represent coordination information and a similar group of classes for versioning, tracking, and managing topics.

Our original GSI topic maps were monolithic files. We soon saw that these files grew to hundreds of thousands of lines, even with only parts of a handful of guides supported. We have accordingly begun to break the topic maps into modules that can be merged according to varying requirements. One module contains the core ontology of typing topics used for both the objects under management and the association roles. A second, much smaller module contains fixed content items, such as classification values. The remaining modules represent the actual data under management. Topics that identify guidance documents occupy at least one module. Each guide will have a set of modules, not all of which are yet fully defined. At this point, we can expect modules for:

- individual text units (*e.g.*, guidance topics)
- associations within a single guide (document and topic hierarchy)
- associations that cross guides (topic derivation or basis)
- associations among content and metadata, including
  - keywords
  - keystones
  - collaboration
  - genealogy
  - rationale
  - background
  - related information
  - usage

Because the metadata elements are of quite varied types, there will probably be more than one metadata module associated with a guide. Keywords are likely to fall into logical patterns with their own hierarchies. Our experience with Ferret led us to topic maps in the first place because we were attempting

to manage complex inferential networks of trigger concepts, which are essentially networks of keywords. While the publishing system may not need such complex networks as the Ferret analytical engine, the pattern established there may nonetheless still be applicable. Collaboration data, as a collection of texts, is quite different from keyword networks and so needs a different kind of storage and a different kind of links into the document from the keywords. As the genealogy has to do with subject identity, it may require an entirely different structure within the topic map. Background information, because it is likely to be in the form of hyperlinks to external data, will probably require a module with proxy topics to represent the external resources and then associations to link the guidance topics and the resources.

In a topic map, topics are proxies for subject matter, not actual repositories of content. In our early topic maps we placed what appeared to be content in topic names (*e.g.*, the numbers and texts of guidance topics both became scoped basename strings) because names are a convenient means of presenting information for browsing the topic map. However, name strings are not appropriate for storing many forms of content, so some appropriate replacement text must be generated for browsable names.

## § Architecture for document management

A conventional technical-publishing system might consist simply of XML editing/publishing software riding on top of a content-management and workflow system that is designed to handle files. Because of the complex web of interdocument linkages and annotation that is faced by ICCP, such a system is only part of a solution. On top of a content-management system there must be another layer of link-management tools that becomes the primary interface for locating and tracking the information that is passed to the publishing tools. One of the most important components of this upper layer of the system is one that analyzes guidance documents into the components to which links can be made or assembles documents for editing from components selected through a search of the knowledge base. Another component is a metadata editor that works in parallel with the document editor. Underneath this layer may be a conventional content-management system, perhaps in conjunction with a database that maintains metadata.

### *Dissector/Assembler*

The central component for making multiple approaches to information work is a software hub that we have called a “Dissector/Assembler.” Although it is actually a suite of programs that can be used in several ways, it is convenient to deal with it as a black box in the center of the system. On one side of this box sits a conventional publishing system that deals with continuous text, whether whole documents or large units such as chapters. On the other side of the box is a collection of software that looks at information in a network view, notably the metadata editor and the topic map tools. Supporting both sides, through the mediation of the black box and under the control of the topic map, is the storage system. The storage system, seen as a whole, must manage both complete documents and document components.

The document view of guidance is, except for the elements that represent the topic hierarchy, typical of the structures found in systems for technical reports. The topic hierarchy consists of nestable blocks of regular structure, such as chapters and sections. Guidance topics are usually collected in special sections called, as might be expected, “Topical Guidance.” An individual guidance topic has a number, the text of the topic, and the classification information associated with the topic. Notes can be added to either the text or the classification. Topics nest directly under the current DTD. Classification information has its own structure, generally consisting of a level (or range of levels), a category or categories, and some subsidiary information for some categories (such as conditions for declassification). Although classification values constitute a controlled vocabulary, the authors prefer to enter them as text, so the DTD supports them as element text. There are tentative provisions for some metadata and for reflecting classification values in attributes.

The dissector component (Figure 1) of the hub takes the continuous document (or document fragment) and breaks it into its components at a granularity appropriate for building the topic map and managing those components under the map’s control.

The dissector works primarily by transformation (Figure 2). As a preliminary step, the dissector currently converts ICCP’s authoring XML into an intermediate streamlined XML that maintains a tree hierarchy similar to that in the source. (This first translation can be reversed by a complementary converter.) The intermediate XML can be used in several ways. Y-12 has an online classification-support system,

Figure 1: Dissector structure

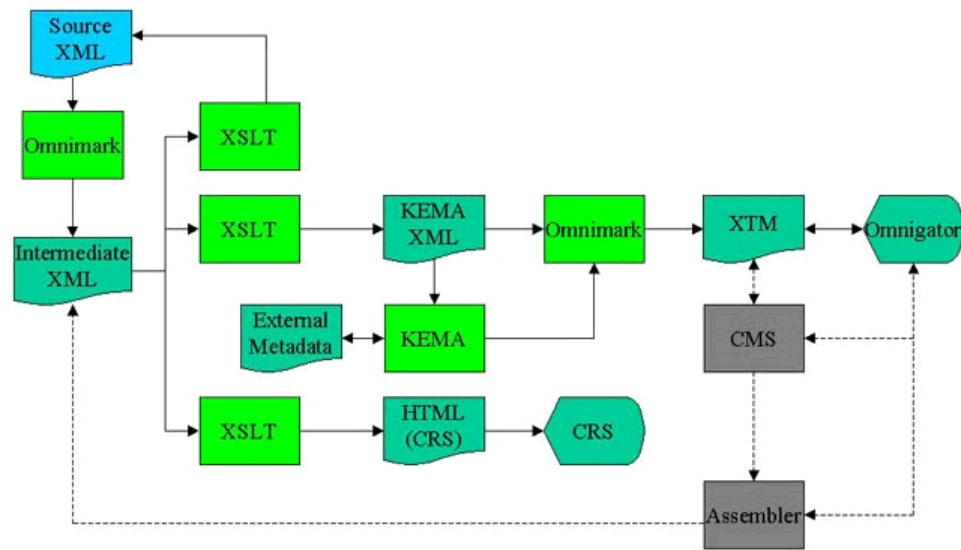
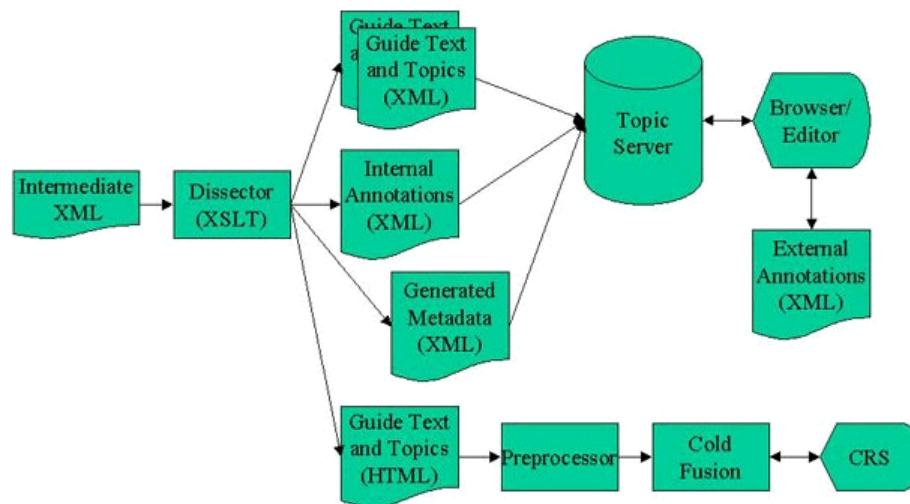


Figure 2: General data flow through dissector system

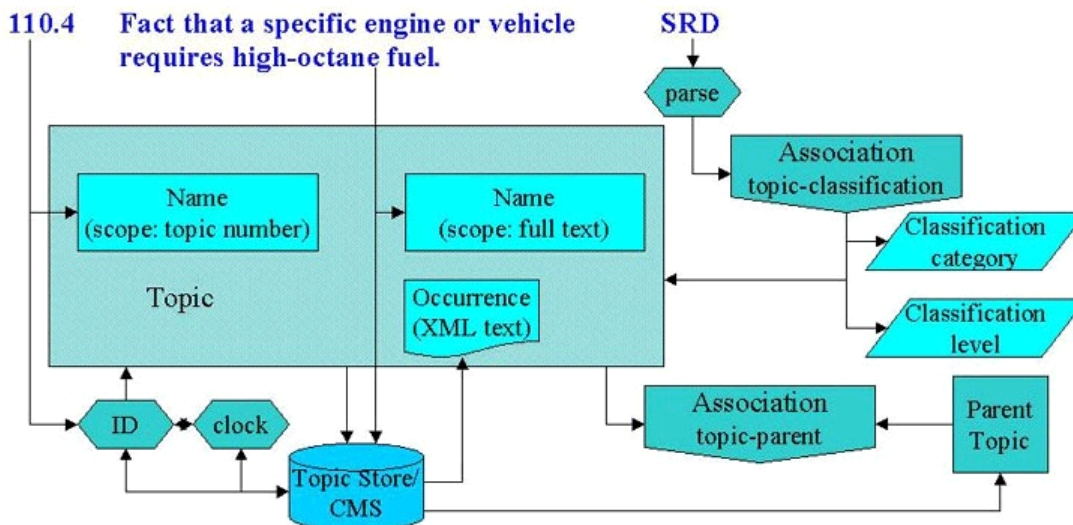


CCRS, and we have a transformation from the intermediate XML to the specialized HTML that is used as input for this system. The transformation of most interest to GSI is into manageable components that will create the topic maps at the center of the guidance-management system.

The first stem in transformation divides a source document into several streams, separating metadata from primary content. The content is then separated into the components familiar to technical publishers, such as sections, headings, and paragraphs, and the content that is unique to the ICCP environment, that is, guidance topics. At this stage of dissection, each managed unit is assigned an identifier that not only establishes its identity in the topic map and content-management system, but also is a key to reassembling documents for the publishing systems.



Figure 3: Dissection of a guidance topic



A guidance topic, with components of topic number, topic text, and topic classification, generates a topic in the topic map (Figure 3). Both the topic number and the text are converted to names, for convenience in searching and display. The text, as text, becomes an occurrence, scoped by the version and date information from the content-management module. Other occurrences may be pointers to XML source files. The classification, as a text string, is parsed and converted to members of an association between the primary topic and topics that represent allowable classification values and related information. When guidance topics are nested, the hierarchy is flattened in the topic structure, but additional associations are constructed to represent the relationships between parent and child topics so that each topic can be presented in context, even within the network of dissected components.

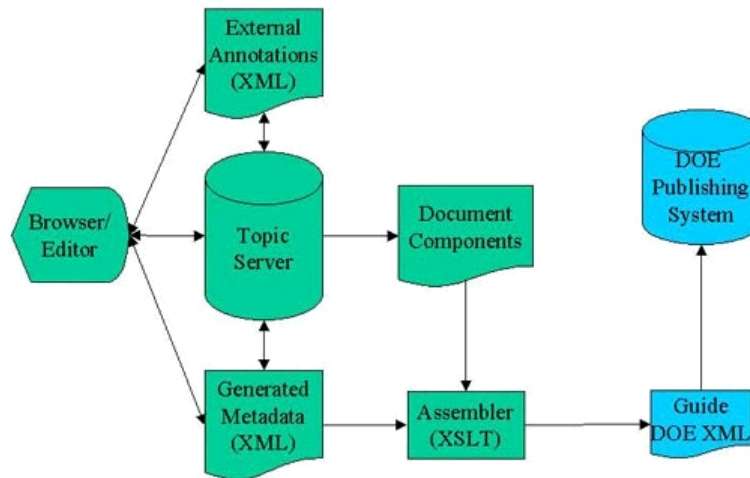
Transformation of topics may require other manipulation of XML in the source document. Dissection is not expected all the way down below some molecular level, such as a paragraph or the major components of a guidance topic. Although most statements of guidance topics are simple text, there is occasional internal markup, such as superscripts in chemical formulae or notes embedded in either the topic text or classification. In such a case, the dissector must do two things. For presentation in the topic-map browser, internal markup must be stripped or disguised because the XTM DTD does not permit markup from foreign namespaces in elements like `baseNameString`. For the actual text to be stored in the content-management module, however, the text must be extracted into a fragment with internal markup intact. Since this text will be manifested in the topic map only as an occurrence link, the foreign markup is permissible.

The assembler component reverses the dissection process (Figure 4). When a target for revision has been selected from the topic map, the assembler will build a suitable document, or document portion, and feed it to the publishing system. The assembler reconstitutes the document hierarchy from association data.

The dissector works with the content-management module of the system to assign and manage identifiers for objects. Identifiers will have both persistent and transitory elements: the ultimate persistent component must deal with subject identity, without respect to current location of the component. At present, the only available components for persistent identifiers depend on the identity of a guide and the topic number within the guide. This is sufficient to identify a guidance topic within a designated version of a guide. However, revision to guides often results in topic renumbering, so document and topic numbers are not sufficient to track a topic through its history. Ultimately, the genealogy metadata will track identities, but a persistent string is needed. How versioning information is presented will depend in part on the content-management system selected to support the project, so details of identifiers are not settled. At the present, the two known elements, current guide and topic numbers, are being used for the demonstration topic maps. Whatever their evolution, these structured



Figure 4: Assembly of components of a guidance document



identifiers are the key to linking within the topic map and also for reconstituting documents from components. The identifiers assigned to guidance components become also the means for attaching further annotations and metadata to the components.

### Metadata editing

We are at an interim state with annotation of documents. Although the current DTD for a guide provides for some annotation elements nested at various levels in a document, as we suggest above, we suspect that is not adequate. We hope eventually to have a topic map editor integrated into our browser, but for the moment, we are using a separate application, KEMA [Knowledge Engineering and Mapping Assistant]. Since KEMA was originally created to support the Ferret application, it already provided for attaching keywords to guidance topics. It can navigate a hierarchy of keywords in the implication trees used by Ferret as well as a collection of guidance topics. We have extended it to allow the attachment of the other kinds of metadata envisioned by the publishing system and GSI.

Our current environment passes a guidance document through the dissector system, from which it can be loaded into both the topic map browser and KEMA. Revisions and annotations executed in KEMA take a different path through the dissector hub and are reloaded into the browser. Updated metadata is then refreshed in the browser. The greatest disadvantage of this system is that it is a batch process, requiring a refresh cycle.

KEMA is a good fit for some relationships within the envisioned metadata, notably the keywords and keystones. It is simple enough to add fields for additional text, such as collaboration. However, KEMA is not a full XML editor and so is not suitable for annotation that requires internal structure or data that involves hyperlinks, such as the background and usage links. For those latter kinds of annotation, we may well have to wait until it is possible to edit within the topic-map browser. The specialized nature of KEMA means that it edits the strings used as names in the topic map, but not the content occurrences, which depend on the XML editors in the publishing system for their creation and revision. As a consequence, the current user interface to the system involves three windows: the publications editor, KEMA, and the topic-map browser, plus the mechanism for running the batch refreshes.

The current dissector system makes it possible to generate a skeleton topic map, containing only the links shown in Figure 3, without intervention. But this topic map is incomplete from the point of view of the whole GSI system because no further metadata, particularly keywords, is present. KEMA allows populating that keyword network, and we are just now beginning to expand the group of KEMA users that is testing the addition of keywords to guides.

## Content management

Content management is simultaneously the simplest and the most complex of the problems facing the system. When the planning for the publication system began, a conventional commercial document-management system was expected. However, once it became evident that conceptual content was the key to managing guidance the requirements for a system became much more complex [ISOGEN 2001]. The DOE-wide approach of GSI, which stretches beyond the central office of ICCP, has only added to the complexity of the requirements. The emerging need for metadata and annotation has increased the requirements even more. To fulfill the vision of GSI, the system must manage concepts as well as documents.

We still expect a conventional document-management system to be a component of the overall architecture because it remains necessary to deal with entire guides, both as XML source and as PDF and HTML output. Conventional workflow is also expected when a guide reaches the final review and release stages. The only complication that may be added to a commercial system may come from the need to connect it to the topic-map interface over the entire system.

Managing content at a molecular, if not atomic, level may also be possible using the same engine that supports the management of whole documents. After all, the basic requirements of check-in, check-out, version control, and object locking may apply equally to both large objects and small. What differs is the requirement to deal with many thousands of objects rather than just a few hundred. If the problems of using a topic-map system to manage the interface to the content-management system for documents are solved, then at least one layer of the problem may be solved across the board.

The major problems with the management of molecular-level objects revolve around object identity. We recognize that a major module in the system, one that must interface with almost all the other modules, will be a mechanism for managing identifiers. In the topic maps that have been developed for GSI so far, we have dealt with guidance documents only in a frozen state. Guidance topics have been identified only by their topic numbers and their parent guides, and identifiers could be generated simply by the dissector scripts without reference to the content-management system. Thawing the documents to support revision immensely complicates the management of identifiers. Identifiers will require versioning components and/or timestamps. Although versioned identifiers are not a new concept, their application to the large number of objects suggests to us that we must have an "ID Server," perhaps created in a database, connected to the system. The project team has already investigated several types of strong identifiers for use in this context.

Management of identifiers is further complicated by the fact that guidance topics are not persistent objects. Over the more than half century of classification within DOE, policy about classification at the highest level has been relatively constant. However, detailed guidance changes for varied reasons (*e.g.*, information becomes known at the unclassified level because of treaty agreements, or it is decided that some information is no longer sensitive). One of the chief reasons for GSI is to rationalize guidance across DOE and eliminate unnecessary redundancies. Some redundancy will probably remain in guidance. At Y-12, for example, some people may need to locate guidance about a material from the perspective of its physical properties, others from the perspective of its use in our products, and still others from the perspective of how it is stored and shipped. Accordingly, a single topic is reiterated with varied language in a guide. The concept of subject identity will simplify maintenance of the multiple avatars of a conceptually unified guidance topic.

We are currently reorganizing some guides to make them more usable. As GSI progresses, we can expect many changes, of which renumbering of topics is one of the simplest. In such a case, the fallback to unified identity becomes essential. Simple identifiers, such as topic numbers, will clearly not be adequate. Topic identification on a conceptual level (through keywords and keystones) is one of the most important emphases of GSI, and it is one of the things that led us to topic maps in the first place. We have long recognized some forms of conceptual identity: for derived guides, such as those issued locally by Y-12 and the other sites, we maintain tables of derivation to trace our topics to those in master guides, even when the wording of the topics is dissimilar. Such tables, of course, are readily transformed to topic maps.

The management of identifiers must thus interact with the management of conceptual identity within the overall system. If we create topics, such as the one shown being created by the dissector, whose names include their texts and numbers (certainly useful for locating the topics for someone familiar with the current guides), we must recognize that occurrences of such topics may be scoped very closely to the versioning information current when they were created. If we supplement such particularized topics with

conceptual locators, we may find that we need to reify associations that bring together the evolution of topics over time.

Conceptual location of guidance topics has further implications for content and document management. Once the topic map is fully populated, we can use association data that involves document hierarchy to reconstitute specific versions of documents from their identified molecular components. We can also work from conceptual data to construct documents that have not heretofore existed, perhaps as the starting points for creating new guides.

## § Future directions

The system we have presented is very much a work-in-progress. The basic publishing system has been tested with actual production of both new and revised guides. We are currently planning the migration of the publishing system into the field sites beyond those involved in testing. The topic-map ontology is stable, and the basic screens for the topic-map browser have been tested. KEMA is being used to evaluate the user interface for metadata editing. The dissector conversion modules are complete, except for the ID server, but the modules are still a series of scripts, without an interface that is convenient for end-users. In the current state, the system is sufficiently complete we can create a guide in XML, process the guide through to a topic map, bring up the guide for annotation, and then view the results in the topic-map browser.

The document assembler is in a more primitive state. At this writing, only one module, which converts the intermediate XML back to the source, is functioning. Completion of the assembler is likely to depend on further work with a document-management system, and we are only beginning to evaluate candidate systems. Which system DOE selects will depend not only on the usual criteria for such systems, but also on the degree to which they can interwork with the topic-map tools.

One of the next stages will be to develop a browser-based metadata editor that is integrated with the topic-map system. Having an integrated editor will simplify the user interface and allow direct transition between searching the collection and annotating it.

Our experience with the pilot projects that have led to this system plan have given us confidence in its success. That XML-based publishing has been successful should surprise no one in the markup community. Our topic maps, though potentially very large and complex, are not untypical of knowledge-management projects. We have already learned from our linking of keywords to guidance topics through topic maps that we can find redundancies in classification guidance. We have learned also that topic-map based conceptual searching provides results that cannot be achieved by other means. Using a concept-based system for locating guidance will be a new experience for the classification community. The only system that has been available to the general community has used conventional full-text searching, with the expected attendant problems. Y-12 began to make an intelligent query system, based on Ferret concept networks, available to a test group last year and now is expanding it to the larger ADC community. With the current ability to generate a skeleton topic map on the fly, we hope to populate more concept networks and test them in the online environment.

Even as we investigate the integration of a document-management system, we are continuing to generate data for inclusion in the system. Several guidance documents have been prepared in XML, and others are on the way. We have keywords assigned to several documents, and more are planned. With the dissector tools we have, we can increase the topic-map corpus already under construction. By the time the system is assembled, we should be ready with enough content to put it into immediate use. We are looking forward to that time.

---

## 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 GSI. In particular, the continued support this initiative receives from the management of DOE offices of Information Classification and Control Policy and Security Policy staff was extremely valuable, in times of scarce resources and competing projects. Many thanks go to the director of the 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), particularly that 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.

Johnnie Grant and Ron Sentell, senior classification guidance authors at ICCP, have contributed to many aspects of the system, from testing the publishing tools and metadata editor, through building the initial paper topic map, to steering the preparation of metadata.

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.
- [**Lê and Mason 2002**] Lê, Vinh, and James David Mason. Topic Maps for Managing Classification Guidance. In *Extreme Markup Languages 2002: Proceedings*. <http://www.idealliance.org/papers/extreme02/html/2002/Mason01/EML2002Mason01-toc.html>.
- [**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 PEPCO [Potomac Electric Power Company]. 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 DOE [Department of Energy], 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 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 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 [Y-12 National Security Complex] in Oak Ridge, Tennessee.

### Extreme Markup Languages 2003

Montréal, Québec, August 4-8, 2003

*This paper was formatted from XML source via XSL  
by Mulberry Technologies, Inc.*