

Retrieving Knowledge In e-Government: The Prospects of Ontology For Regulatory Domain Record Keeping Systems

Richard J. Daddieco, Ph.D. Candidate

School of Information Science, University of Pittsburgh, 15260
rjd7@pitt.edu

Abstract. If e-Government systems are to achieve the anticipated goals of storing, manipulating, diffusing, and preserving knowledge, then these systems must incorporate mechanisms for domain-specific information. One often overlooked asset for this activity is the knowledge embedded within and between records and legacy record keeping systems; however, lacking government-wide documentation management systems and subject specific resource description, timely retrieval of relevant records is impossible. This paper explores the benefit of ontologies for improved retrieval of subject domain knowledge. Historical records and recording keeping systems were reviewed for this purpose. Structural and semantic information was derived, from which requirements for an ontology in the subject domain of export control regulation were identified. Class, subclass, instance, and relationships were identified and modelled via an Entity Relationship Diagram. Through this exercise, it is clear that e-Government systems hold great potential for circumventing past information management deficiencies, and ontologies may have a role in the transition from information systems to knowledge systems.

1 Introduction

Just as developers are rapidly developing and deploying e-Government systems, so too is considerable attention centering on precisely how effectively these systems store, manipulate, diffuse, and preserve knowledge. The failure of countless technology ventures over the last decade clearly demonstrated that blind application of technology constitutes neither a necessary nor sufficient condition for effective knowledge management. Systems developers must do their best to incorporate as many existing knowledge artifacts from the policy community they seek to connect. This explicit goal must precede the application of technology, and in no place is this fact more germane than in the systems development process of e-Government.

It is as yet unclear how successfully e-government systems are being developed. This is not to suggest that the prospects for e-Government are gloomy; in fact the possibilities are limitless. However, numerous examples of information failure within government foreshadow the prospects of e-Government. The

concern is whether e-Government systems can overcome the systemic, organizational, and bureaucratic barriers to effective knowledge management or instead simply replicate the latent deficiencies so long present in analog systems. Information failures are costly; however, in instances of government decision-making these costs can prove incalculable.

One area that has long suffered from knowledge mishandling is the administrative process of regulating militarily sensitive exports or the area of export controls. Instances of information failure in this area are frequent, and they characterize the problem plaguing knowledge management in governance more broadly, that of knowledge retrieval and sharing. The case of the *U.S. vs. Fiber Materials, Inc.* is a prime example.^[1]

In 1995, two businessmen were found guilty of violating U.S. export control laws based on the memory of two government officials as to the meaning of two words in a specific export regulation: “specially designed.” The two licensing officers testified at the trial that the meaning of the terms had never been written down. It was “in their heads.” Researchers for the Defendants in the case, however, successfully unearthed the explicit, written definition of the terms as it was intended to be used in the regulatory language. The very government agency responsible for issuing and enforcing the regulations, the same that had that brought the charges, had defined “specially designed” some forty years prior to the trial. The definition had simply been lost, encapsulated within an obscure government document.

Because the written definition contradicted government testimony, in August 2003, the original verdict was overturned. This knowledge failure calls into question the validity of every export license decision and compliance with regulations which included the terms “specially designed” prior to the discovery of the documentary evidence. The terms appear in over two-hundred export control regulations. Did government officials deny licenses for benign, profit generating exports? Did they accidentally permit the release of strategic items, proliferating weapons technology. Lastly, why was this crucial knowledge, so important for export license decisions, seemingly lost for over forty years?

e-Government systems must push information into knowledge. Systems must operate within well-defined and informed parameters of specific subject areas. This knowledge must be bound. It must include the language and process derived from accumulated, relevant evidence recording administrative history, such as documents. Ontologies may prove a useful tool for facilitating there retrieval.

2 The Study

This paper endeavors to build an ontology for the subject domain of export controls. Development of such an ontology may prove critical for effective knowledge retrieval and knowledge sharing in this area. Via detailed analysis of public records associated with export controls, this paper constitutes a first iteration in this endeavor. It is in no way to be viewed as complete. Ontology building is an iterative process. The intent of this first iteration is to examine the many lay-

ers of relationships between information objects, and, in so doing, inform future development through a document-centered approach. First level taxonomies can guide ontologies. This analysis revealed a central taxonomy common to multiple objects around which an ontology could be built.

This document analysis relies on international and U.S. Executive branch records obtained from the U.S. National Archives, U.S. government agency libraries, three U.S. Presidential Libraries, multiple depository, public, and academic libraries. A representative sample of the documents reviewed include:

- Declassified documents of NATO’s Coordinating Committee on Multilateral Export Controls;
- Declassified Executive branch documents obtained from Presidential Libraries;
- Title 15, Part 799 U.S. Code of Federal Regulations: Commodity Control List, copies of which were reviewed between the periods of 1950 and 2002;
- Schedule B: Statistical Classification of Domestic and Foreign Commodities Exported from the United States;
- U.S. Department of Commerce Annual Reports;
- Department of Defense “Militarily Critical Technology List;”

In total, over twenty libraries, records, and information centers were scoured for information relevant to the domain of export controls.

2.1 Domain, Memory, & Decision

The movement of information within the bureaucracy of arms control and non-proliferation policy is extremely complex. It is a policy area that involves dozens of countries all of which are members of several international organizations and regimes. Representatives from member countries negotiate agreements. Each of which results in national legislation and the promulgation regulations. The domestic policy and regulatory process is equally complex.

Each national government must implement international agreements at home. This typically involves regulating the movement of goods and services across national borders. This is an administrative function encompassing the recording of shipments, liaising with the public, enactment and enforcement of regulations, the licensing of exports, and collaboration with multiple government agencies including intelligence, crime enforcement, financial, and diplomatic staffs. Within each of these government agencies exist specific administrative operation. Moreover, inter/interagency teams further subdivide bureaucracies, often requiring their own administrations and procedures.

Within this administrative spider’s web sits the decision-maker. Decision makers often share their experience, explaining what may have been done in similar situations. However as stated earlier, the export control process is complex. It is composed of multiple bureaucracies and agreements. The decision-maker may need information about the process and language of regulating, which may reside outside their bureau or immediate functional domain. This policy context is composed of structural, sequential, and linguistic information or taxonomies.

Each adds evidential value to the chain of administrative process. The decision-maker can rely on this context frame for decision-making, provided it is modelled and retrievable.

In this process of classification, precise language is critical for assigning meaning to the export reflected in the form of the export license request. But regulatory language is widely charged as terribly vague, ambiguous, and unclear. Unfortunately, export licensing officers must render decisions in this ambiguous environment, and evidence on which to base his or her judgment and perpetuate the spirit of nonproliferation policy is unclear, as the aforementioned case suggests.

2.2 Ontology Defined

What is an ontology? Jasper and Uschold [2] define the primary purpose behind ontology as “reducing the cost of multiple applications having common access to data.” This may in turn, facilitate inter-operability. Lavery & Boldyreff [3] offer a more refined definition, “a domain specific ontology is a knowledge management tool used to support communication and knowledge reuse about a specific domain.” This interplay between place and time—that of centralizing knowledge and communication, is the heart of ontology. For the purposes of this paper, ontology is best defined in the words of Noy and McGinness [4] as “a formal explicit description of concepts in a domain of discourse.” Conceptually, it is useful to think of an ontology as similar to an office filing system.

Filing systems support an organization’s information needs in that they provide common access to records for a variety of organizational objectives. Customer files, employee files, product information, competitor information is centrally organized and stored for future use. The filing system structures and orders information, knowledge typically stored in the form of paper records but now including both paper and digital objects. This system utilizes the contextual language of the enterprise’s operational vernacular or what information organization researchers have characterized as its “controlled vocabulary.” [5]

This reuse of information is particularly important for government decision-making and should inform the application layer of information systems and metadata resource description. Ontology has replaced the filing cabinet in the ubiquitous platform of the World Wide Web, while the filing system has been replaced by metadata description frameworks. Ontologies may connect information objects, rebuilding semantic and organizational knowledge. This benefit strengthens ontology’s resonance as a method for developing description frameworks. Requirements engineering methods are critical for ontology building; however, one of the key inputs of requirements engineering is frequently dismissed: policy records.

3 Rudimentary Structure Revealed Through Documentation

Records professionals are cited as being in a strategic position to influence the information architecture of information systems. [6] Toward this end, scholarship has centered on the elicitation of requirements for documentation systems and the evidential value of records.[7–10] Records can play such a substantive role in eliciting the nature of information objects across such enumerative categories as: taxonomy, administrative and organizational structure, and function. In the export control domain, these three aspects are intricately interconnected, woven together via a complex community of subject area specialists. These relationships between actors, sub actors and the language through which they communicate is embedded in historical records and government publications. Through their analysis, the beginnings of an ontology may come forth.

4 Fields & Description

Content analysis revealed definitive, stable knowledge structures within, between, and across documents. Most contained an area heuristically defined herein as a “header section.” Within this header, numerous fields contained specific and distinct types of information, such as date, document number, and title. When placed with the context of the taxonomy of the policy domain, these fields also depict and described such things as transaction, organization, function, custody, and interdependency. Each of these resultant categories of information also serve as information objects and classes, the backbone of the subsequent ontology. These objects contained within records inform the reader of the context of the documents origin and policy domain knowledge.

5 Taxonomies: Semantic, Structural, Functional

Surprisingly, the structure of the CoCom records remained constant over time. The relevant information on which negotiators, domestic policy analysts, and regulatory staff relied did not change, a particularly telling feature when compared against the various action areas associated with the generation and handling of the records. These action areas correspond to structural and semantic taxonomies and process of policy administration. Within this process resides the agents and language of the domain.

5.1 Taxonomy to Class

When records were compared against the structure of export control regulations, shared classes and a common taxonomy between the structure of CoCom’s record keeping system, the structure of the CoCom records themselves and the structure of the organizational system in the domestic regulations surfaced. This taxonomy linked multiple documents in this policy domain.

Given this relationship, it is clear that a method was in place that would inform policy officials of the precise negotiating history of the specific goods and technologies under the international body of CoCom. Domestic policy analysts and regulators could have relied on the evidentiary value these records and record keeping systems provided, had systems been in place that enabled the retrieval of this knowledge.

Content analysis revealed multiple sub-taxonomies as well. Free floating subdivisions within the body of the Export Administration Regulations, i.e. “parts, components, and accessories” was also tied to various document titles in a variety of different organizations. Export Control Commodity Number is a unique identifying number derived from multiple indices. The categories in these indices reflect commodities, country, the international organization or policy associated with the regulation. These various categories also reflect action areas executed in multiple offices and areas in the administrative chain.

ECCNs form a semantic taxonomy used in multiple classes derived from shared technical or functional qualities between commodities. For example, metalworking machinery is a class identified with the digit 2 in an ECCN. Similarly, parts and components is a subclass in the ECCN which is identified by the letter B. Several categories or facets are represented via the sequential stringing of digits. Similarly, each digit represents a specific office, function, and language in the administrative policy process.

Surprisingly, though the digits representing various classes may have changed over time, the class names remained fairly constant. These classes reflect strategic technology areas. One can trace the controls on a specific technology area over time through the retrieval of documents commonly classified as germane to this technology area.

6 The Ontology Form

The form of the ontology is represented in the corresponding entity relationship diagram. The entities identified in the chart should be read as principle agents, each of which display characteristics or attributes. In the relationships between classes, many subclasses result. Some of which appear as “free floating” in that they may reside within any number of superclasses. This “class hierarchy” is represented within the ontology in the various relationships between the classes and attributes, hence enabling identification, collocation, and retrieval of knowledge.

7 Class, Subclass, & Relationships

Functional and organizational interdependencies were identified across export control records. The following entities and relationships describe the conceptual pattern of interaction between taxonomy, structure, and function. According to Dragomir Radev (2000), the possibility exists for further differentiation by relationship type.[11] From the multiple layers and inter/intra document connections the following entities and relationships were identified:

- **Commodity Control List:** The U.S. list of regulations devoted to “dual use” goods and technologies. It was derived from the Industrial list of NATO’s Coordinating Committee. The CCL *consists of* ECCNs.
- **Export Control Commodity Number:** Derived from several subject indices and it is used as a numerical identifier in the U.S. export control regulations. *ECCN encodes* Commodity Category, Commodity Group, Type of Control, Unilateral or Multilateral, Reason for Control (1-to-m).
- **Negotiation Transcript:** The record of negotiation *Negotiation Transcript identifies* ECCN(m-to-m)
- **Reason For Control:** The reason for control is a nominal policy category which may or may not restrict an export from leaving the United States. **Subclasses** consist of regime, Agency, Regulatory Unit, Negotiated Transcript, Regulation.
- **Agency:** A government organization chartered with the responsibility of some aspect of policy and/or administration. The specific bureaucratic body that is responsible for the administration of the regulations. Agency consists of Administrators, Bureaus, Departments, Working Groups. *Agency creates* Negotiated Transcript (m-to-m).
- **Regime:** A specific, institutionalized understanding dedicated to a common arms restriction goal. Regime consists of Country, agency. *Regime creates* Negotiation Transcript (1-to-m)

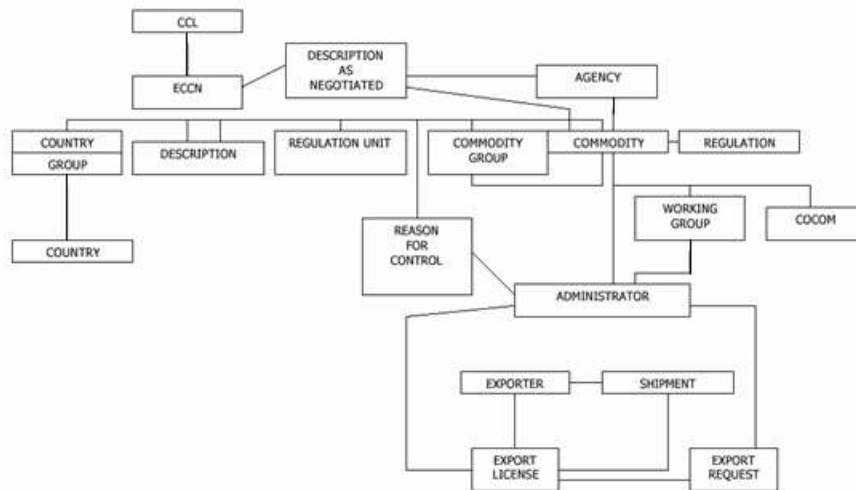


Fig. 1. Export Control Entity-Relationship Diagram

7.1 Relationships Between Documents And Organizations

Further analysis of CoCom's records when compared against the structure of the Export Administration regulations reveals unions or intersects of classes of information. This intersection or overlap of classes between these two organizational systems denotes relationships between the two. CoCom records informed decision-making domestically in U.S. bureaucracies. History provides direct lineage between the negotiations in Paris and the subsequent passage of legislation and regulations promulgated under the acts. The tacit relationship between information classes and the benefits it may provide for current decision makers is illustrated via document comparisons. The composition of CoCom document number reflects the descriptive categories contained in the Export Administration regulations. This shared information is aptly defined as class in ontology terms. Relationships between these information objects and their class hierarchy can then be traced and conceptually unified to enable users to retrieve an array information objects which may inform decision making. Because these objects are interrelated, each informing the context of the other, both would provide historical content for a decision, context which can prove crucial for consistent decision making on export licenses over time, context which might inform future policy.

8 Resource Description: XML & DTD

With the conceptual framework defined, resource description may include definitive class and class hierarchies and associations to related classes. These descriptive patterns may be reflected in extensible markup languages such as XML, structuring these classes, hierarchies and associations within document type definitions. A much more exhaustive analysis should precede the development of such items, more exhaustive than the rudimentary analysis reflected in this paper. Ideally this analysis would include more than analysis of content in historical documents, but also substantive interviews with members of the user community. Unfortunately, such developments are beyond the scope and resources of this study.

9 The Benefits: Knowledge Centralization, Sharing, & Preservation

As mentioned at the outset of this investigation, one of the primary benefits of developing such ontologies for E-government systems is centralize knowledge on subject domains. With the structure outlined above, we see an ontology in its nascent stage. The benefits of which include:

- The reuse of critical knowledge;
- the sharing of domain knowledge across a variety of settings, including within organizations, across organizations nationally, and within communities of practice that extend across borders.

- the preservation of knowledge.

As technology advances, so too will its military significance and the aforementioned classes. These changes may be tracked via conceptual concordances. Preserving this historical lineage and context within the widely dispersed, subject specific export control community.

For example, the changes in militarily sensitive technologies are reflected via different commodity groups between 1991 and 2001 as seen in table 1.

Commodity Groups (2002)	Commodity Groups (1991)
0-Nuclear Materials, Facilities and Equipment	0-Metal-Working Machinery
1-Materials, Chemicals, Microorganisms, and Toxins	1-Chemical and Petroleum Equipment
2-Materials Processing	2-Electrical and Power-Generating Equipment
3-Electronics	3-General Industrial Equipment
4-Computers	4-Electronics and Precision Instruments
5-Telecommunications and Information Security	5-Metals, Minerals, and their Manufactures
6-Lasers and Sensors	6-Chemicals, Metalloids, Petroleum Products, and Related Items
7-Navigation and Avionics	7-Rubber and Rubber Products
8-Marine	8-
9-Propulsion Systems, Space Vehicles and Related Equipment	9-Miscellaneous

Table 1. Commodity Group Comparison

One could trace the change in number between categories and the corresponding structural changes, say the transfer of one function from one office to another and the corresponding linguistic differences as evidenced in that office’s documentation.

10 Implications

What then does this mean in the context of export license decisions? Primary methods for determining prior action on export license requests require searching of database/s of previous commodity classification decisions. These commodity classifications reveal prior licensing decisions, but they do not reflect subject domain knowledge. The database communicates little to no historical information in terms of how to interpret regulatory language. Documents recording the policy discussions of the formation of export controls would prove extremely useful in this situation, and document management systems may facilitate their retrieval. The formal logic outlined in the preceding ontology would prove helpful for this

task. It would allow for narrowed searching of relevant records and enhanced the subsequent use of natural language search techniques.

With the above ontology embedded within a Web-based system, user's could search for documents associated with specific commodity categories and policies, emanating from different bureaucratic organizations, different staff or administrators, or even different international regimes. User's could share a common, administrative, knowledge base. For example, a user could query for "all documents" with reason for control "nuclear." In a database, the query would take the following form:

```
SELECT Document.identifier FROM Document Descr, ECCN, RFC
WHERE Doc.ECCN = ECCN.ID AND
ECCN.RFC = RFC.id AND
RFC.description = "nuclear"
```

Through resource description standards for the subject domain in the way of custom metadata schemas, descriptive classes would allow for narrowed recall of documents. From there NLP could further refine searches, based upon keyword matching in the querying of the full text of the documents . In both instances, only those documents that discuss nuclear related goods and technology would be retrieved. Policy officials could then rely on the decisions of their predecessors, the negotiations of the international organizations within which these regulations were negotiated, and moreover, memory could extend the walls of their respective organizations. We would see the beginnings of policy memory.

11 Conclusion

This investigation scratches the surface of the potential for ontologies in e-Government Web-based systems. Developers must take into account the specific subject domain as well as the many layers of administration if they are to facilitate resource sharing. At the same time, however, development must transcend the methods of traditional systems building, and include at the forefront of the design the knowledge and needs of the user and user community. Ontology development is indeed a powerful response to that need.

References

1. Daddieco, R.J.: US vs. FMI: A case study in knowledge failure in export controls. In: this volume. (2004)
2. Jasper, R., Uschold, M.: A framework for understanding and classifying ontology applications. Technical report, (Boeing Math and Computer Technology)
3. Lavery, J., Bolydreff, C.: Developing an ontology. Technical report, University of Durham (2002)
4. Noy, N.F., McGuinness, D.L.: What is an ontology and why we need it. Technical report, (Stanford, CA)

5. Taylor, A.: *The Organization of Information*. Libraries Unlimited, Inc., Englewood, CO (2002)
6. Robins, D.: *Information architecture, organizations, and records management*. *Records and Information Management Report* **17** (2001)
7. Bearman, D.: *Constructing a methodology for evidence*. In Bearman, D., ed.: *Electronic Evidence*. (1994) 1–18
8. Bearman, D.: *Electronic evidence: Strategies for managing records in contemporary organizations*. *Archives and Museum Informatics* (1994) 360pp.
9. : *International research on permanent records in electronic systems (InterPARES) project* (1999)
10. Bearman, D., Sochats, K.: *Formalizing functional requirements for recordkeeping*. Technical Report Research Report LIS055/LS94001, University of Pittsburgh (1994)
11. Radev, D.: *A common theory of information fusion from multiple text sources, step one: Cross-document structure*. In: *1st ACLSIGDIAL Workshop on Discourse and Dialogue*, Hong Kong. (2000)