Integrating knowledge management tools for government information

Georgia Prokopiadou, Christos Papatheodorou and Dionysis Moschopoulos

Department of Archive and Library Sciences, Ionian University, Palea Anaktora, GR-49100, Corfu, Greece

Available online 5 March 2004.

Abstract

Government information is considered to be essential for any organization's business transactions, decision making, or information rendering to citizens. However, the public sector lacks information units based on the advanced information tools and standards needed to facilitate efficient information management, flow, and exchange. Furthermore, government information is fragmented and dispersed due to (a) legislative and administrative diversity, (b) a complicated administrative hierarchy, and (c) discrepancies in policy implementation regarding the accession and exploitation of information at central, regional, and local levels. This article introduces a digital library architecture for the management and delivery of information either produced or disseminated via public services. Ontologies, taxonomies, and thesauri provide for thematic, geospatial, and administrative hierarchy representation and navigation and ensure an advanced and high-quality knowledge-based framework for information management, search, and retrieval. International standards and formats set the basis for interoperability in the midst of legislative, administrative, and geospatial information diversity.

Article Outline

1. Introduction

The public sector produces and disseminates huge amounts of information, either concerning legislation and its modifications, or governmental operations, or citizen–state transactions. This information either has the form of legal acts published in an official gazette, information sheets/brochures that are published for the assistance of citizens in their transactions with public services, or public records created, circulated, and maintained by public services as evidence, documentation, and information concerning their activities and transactions. Central or regional governments are responsible for the production and dissemination of such information. However, the information flow procedure is considered to be overcomplicated due to the multilevel and multidimensional administrative hierarchy of the public sector. More specifically, national governments consist of collegial government organs, ministries, public organizations, and independent administration authorities. Each ministry is divided into general directorates, directorates, departments, and offices, as well as corporate bodies and committees. Public organizations are separated into legal entities under public law, legal entities under private law, state companies, and mixed economy enterprises. Regional governments consist of regions, prefectures, and municipalities. All these public services have their own goals and competencies to attain. Within this framework, there are certain posts (positions) which themselves have their own hierarchy.

An upgrading of available information systems is called for, as government information:[1. and 2.]

• Provides authentic and reliable evidence of any organization's business transactions. Indeed, greater efficiency in information on legislation and record keeping is the key to unleashing the full benefit and potential of knowledge management.
• Facilitates the decision-making procedure, as it maintains a collective memory of the organization's dealings.
• Assists citizen transactions while respecting the right to freedom of information.
• Contributes to economic development and the proper function of internal markets, as no business activity can be performed without accurate information on the relevant domains. The Green Paper declares that "economic actors cannot make fully informed decisions" without being properly informed.
• Constitutes an education system, as students can learn of the history and the context of public administration issues from objective and reliable sources.

Hence, government information is fragmented and dispersed due to legislative and administrative diversity; a complicated administrative hierarchy; and discrepancies in policy implementation regarding the accession and exploitation of information at national, regional, and local levels. Furthermore, public services lack organized information and knowledge management systems necessary to implement international standards and formats. Thus, the public sector is unable to ensure data interoperability, efficient information delivery, and well-organized information flow within multilevel administrative channels. In consequence, these factors result in a "lack of transparency for citizens, employees, and administrations at all levels" of the public sector.[2.]

Current government information needs call for information systems and services able to facilitate the organization and delivery of government information and ensure accurate and immediate search and retrieval of the required information. Such systems should provide information retrieval, data exchange, metadata homogeneity, and proper information dissemination through administrative channels of national, regional, and local governments.

Recent initiatives do not entirely meet the information needs of the public sector. There are four types of automated services that exist for legal and government information rendering (a) bibliographic databases, (b) full text databases, (c) Internet portals, and (d) information and knowledge management systems. The existing bibliographic databases offer only a description of the required information. As such, the user does not obtain the specific information they are looking for, but only a description of its source location. On the other hand, full text databases present the material and locate the information needed. However, the search engine of such systems has certain drawbacks, such as limited searchable indexes (in most cases, full text databases only maintain one searchable index, namely, "keyword"). Hence, full text databases are characterized by the difficulty of attaining a high level of conceptual search and retrieval.[3.] Portals provide only for the identification of sources that may host/obtain the searchable information. [4.] However, Web searching does not ensure high quality and relevant information retrieval due to heterogeneous information located at different sources. [5.] Moreover, it is time consuming for the users to reexecute their query until the desired information is located. Finally, the information and knowledge management systems either do not cover the available government information in its entirety, as they only focus on specific types of government information, for example, legislation or public records. [6., 7., 8., 9., 10. and 11.] Or these systems do not facilitate advanced specifications for both data and metadata structure, as well as for information retrieval.

This article introduces a digital library architecture capable of providing user-friendly, high-quality, and efficient retrieval of the available government information according to international standards and formats. The proposed architecture aims to code the administrative, geographic, and thematic hierarchy of the public sector by creating and using advanced knowledge management tools such as taxonomies and thesauri. These tools enrich the semantics of the metadata structure and facilitate the information flow within multifaceted administrative channels. The next section presents the proposed architecture, while the following section examines the metadata syntax issues and the existing description standards. Subsequently, the article analyzes the semantics management subsystem of the architecture and presents how one can encode and describe the thematic, geographic, and administrative hierarchy in government information. Next, the article presents a partial implementation of the suggested model and discusses selected works. Finally, the article offers conclusions and recommendations for future work.

2. Architecture overview

2.1. Content, metadata, and retrieval

The proposed architecture of Fig. 1 consists of a query planner via which the end-user attains access to governmental information. The query is set to a user interface either in the form of terms or in plain text. The query planner, responsible for efficiently handling the query by collaborating with the proper components of the digital library, handles the submitted

question and proceeds by deciding where the query must be executed, for what it must search, and which of the results must be returned.

The architecture relies on a pool of digitized administrative information, which is the content of the database, and consists of the following:

• Legislation: this category includes all types of legal actions (i.e., laws, circular letters, announcements, etc.) published in the Official Gazette.
• Information sheets: brochures that are published for the assistance of citizens in their transactions with public services (e.g., a pamphlet providing information on the standard procedure for acquiring a birth certificate).
• Public records: material that is "created, received, and maintained as evidence and information by public services in pursuance of legal obligations or in the transaction of business." Public records may be the subject of official or personal interest. For example, the procedure for issuing a social security number may have official importance. However, the issuing of a license for business operation has personal interest.

The above mentioned categorization is influenced by certain parameters, such as information rights and acquisition demands. The first two categories are utterly accessible by all citizens. However, access to the last category of government information is based on the target audience. Three types of access rights might be introduced reflecting the information needs of each audience category: (1) records accessed by all citizens; (2) records accessed by public servants of different services; and (3) records accessed only by employees of a particular public service.

Metadata of the available Public Sector Information (PSI) are produced to ensure efficient information search and retrieval; government information delivery (through the identification of available material in the public sector); the recording of government information so that the information can be described, located, managed, updated, and hence used more efficiently; reliability in bibliographic and thematic description; interoperability in systems and standards; and policy compliance.[12.] Structure in data is an "important requirement for machine-processable information." [13.] The proposed architecture produces two types of metadata:

1. Metadata for bibliographic description, which refer to all information needed to uniquely identify an item (e.g., title, statement of responsibility, publication and production data, section titles, associated records, etc.); and
2. Metadata for thematic description, which refer to public authorities (services), public authority competencies, public servants, posts within services,[14.] geographic names, and ranks. Thematically describing and encoding government information simply reflects the need to represent a thematic and administrative hierarchy.

2.2. Knowledge management

The (index) terms used to describe an item (e.g., a circular document) are derived from the digitized documents. Index terms are categorized into topical terms, geospatial names, corporate bodies, and personal names. Topical terms refer to competencies, posts, and ranks. These terms are conceptually interlinked via thesaurus-like associations, forming a

multidimensional semantic network for the thematic linking of the entities and concepts used to describe governmental information. The semantic network, besides the topical subjects, depicts and associates the established governments (corporate bodies) and their competencies, with their geographical distribution, their posts, and their staff. Although thesauri usually represent concepts based on topical subjects, in the case of government information, thesauri also need to cover corporate bodies, geographic names, and personal names. The thesaurus standard allows proper names to be included in a thesaurus of subject (topical) descriptors. In this way, names of persons, corporate bodies, and places function within the same semantic pool as topical terms. The interlinking of all the above is necessary to represent administrative functions.

The administrative hierarchy of the public sector may be identified by filtering the nodes and edges of the semantic network. Thus, a taxonomy is built depicting the public administration structure by introducing multiple hierarchy levels (classes and subclasses) for the existing objects (such as ministries, directorates, etc.). The given classes are completely defined by specifying their properties and their restrictions on the property values.

The navigation through the nodes and edges of the semantic network provides meaningful information concerning the public administration regulations, functions, transactions, and its geographical distribution, while the navigation through the taxonomy provides useful information on the government structure. Since the taxonomy results from the semantic network, clearly the two knowledge representation and management tools can interact and complement each other to provide an ontology for all government information. The derived terms (concepts) are interlinked thematically and geographically via the thesaurus-like semantic network and administratively according to taxonomy principles.

Hence, if a minister is interested in examining the number of educational units currently running in the Attica prefecture to define the parameters for economic analysis, the system, via the thesaurus-like semantic network, would present the following information for each educational unit: (1) the type (whether a primary or secondary school); (2) the location (to identify whether the unit functions in a developed or underdeveloped municipality/area); (3) the posts of the school that are currently occupied; (4) the number of employees that constitute the school personnel; and (5) the specialization of employees, as well as their ranks in the public hierarchy. Moreover, if a citizen is interested in getting a driving license, the system, via the taxonomy, will not only provide information about the required procedure for getting the license (e.g., all the necessary certificates that must be obtained), but also suggest the section of the public sector that is responsible for its issue.

3. Syntax issues and description standards

The metadata structure reflects the need to decide and select a specific standard to secure interoperability. However, the choice of an encoding and description standard is influenced by several parameters, but mainly by the type of the material described. As mentioned above, government information is categorized into three areas: legislation, information sheets, and public records. All three types have their own requirements and description peculiarities. Therefore, each of them must be studied as a separate entity.

The first type of government information is legislation, which is published in the Official Gazette of each nation. The latter is a serial publication with an irregular publication

frequency. Public access to legal acts may be ensured by the analytic cataloguing of legislative texts. Universal Machine Readable Cataloguing (Unimarc) format,[15.] which is an implementation of the ISO 2709 standard for bibliographic description, utterly meets the information description needs of this type of publication. This assertion is reinforced by the results of an ongoing project, currently running in Greece. Describing legislation with the Unimarc format has proved that the ISO 2709 standard meets the information needs of legislative actions by providing fields that correspond to a proper information structure. [16.] Moreover, MARC formats (Unimarc, Usmarc, Ukmarc) are used in most of the bibliographic agencies in Europe and United States, including libraries, which makes ISO 2709 the most widely used standard for bibliographic encoding.

Unimarc may be also implemented for the description of information sheets. More specifically, the encoding of brochures, leaflets, pamphlets, and other printed material with informative content could be formed into groups. This would allow such material to follow the rules for the description of collections created by the cataloguer.

The encoding and description of public records are significantly distinct from the other categories. Regarding public record description, ISO 9000 mentions the need for an organization to hold "quality records," reflecting the need for quality and proper record management. However, no reference is made to the methods, techniques, and functions that are required to produce and maintain "quality records." ISO 15489 was designed to meet this need by providing guidance, instruction, and principles for quality record management.[17.] In particular, ISO 15489 defines the technique and the required parameters for the management of administrative records while it also sets the criteria for efficient control of the "creation, receipt, maintenance, use, and disposition of records, including processes for capturing and maintaining evidence of and information about business activities and transactions in the form of records." ISO 15489 is a consequence of internationalization as it derives from the Australian standard AS 4390. However, ISO 15489 does not refer to the management of records that are part of archival collections, even if the records are characterized as administrative. Our focus is on records that are created for business transactions within organizations.

Several metadata standards exist that may be implemented for the metadata encoding of public records, such as Dublin Core[18.] and Encoded Archival Description (EAD). [19.] However, these standards cannot totally meet the information needs of government information.

Public records are characterized by the necessity of describing and encoding not only the item itself but also the broader "collection-thematic or administrative" to which it belongs. Public administration is ad hoc hierarchical. Therefore, the need to map administrative, geographic, and thematic hierarchies of the item being described is most typical of the public sector. However, the existing standards for item and collection indexing and encoding (MARC and EAD for item and collection description, respectively) do not permit the description of the item along with the collection it belongs to. The MARC format focuses only on item description. However, it does allow the identification of hierarchical links at set and subset levels between two or more items (bibliographic records). This interlinking constitutes just one hierarchical level (above or below). Therefore, the need for representing iterative hierarchical levels cannot be met. EAD provides only for collection description and encoding and not for the items that compose this collection.

Within the same framework, the Dublin Core provides for an elementary description of the required information, but unfortunately the offered metafields cannot describe the information in its entirety, as only a fifteen-element set of descriptors are offered. Hence, many countries have proceeded with an extension of the Dublin Core standard. Such cases are as follows: (a) Government Information Locator Service (GILS) for United States,[20.] (b) e-GMS [21.] (UK Government Metadata Standard) included e-GIF [22.] (UK Government Interoperability Framework) for United Kingdom, (c) Australian Government Locator Service (AGLS) [23.] Metadata Standard for Australia, and (d) New Zealand Government Locator Service (NZGLS) [24.] Metadata Standard for New Zealand.

The initiative of the Dublin Core extension is part of the 2004 target for electronic government, whereby all public services will be utterly automated and interlinked, based on a common policy for record management and common standards for metadata description and encoding. Within this framework, many activities have been initiated, such as the development of a metadata standard for geographic information. A good example is the "Content Standard for Digital Geospatial Metadata,"[12.] which is derived from other international standards such as ANSI X3.51-1975, ANSI X3.30-1985, ANSI X3.43-1986, ANSI X3.61-1986, and ANSI X3.172-1990. [25.] It is also worth mentioning the ISO 19115 standard, which refers to the description and encoding of metadata for geospatial information. Moreover, the "Intra-Governmental Group on Geographic Information," residing in the United Kingdom, is actively engaged in offering guidance to organizations regarding the encoding, delivery, and exchange of geospatial information in a range of public records.

Evidently, current initiatives involve the expansion of available standards to meet the information needs of public records. Such attempts are crucial to the evolution of interoperability and system compatibility, even though manuals on the mapping of similarities and differences between these metadata standards do exist.[26.] Indeed, it is necessary to ensure a proper environment for high-quality and accurate metadata exchange. One solution to the problem may be the adoption of both standards under the same platform (i.e., XML). After all, a framework has already been developed that enables MARC to function under an XML environment. [27.] EAD has also been structured so as to work under XML. Therefore, the need for interoperability between different metadata schemes still exists. Recent technological developments have produced a DTD that meets this need. By integrating different standards, it may be developed for the description and encoding of existing government information. However, such research is not within the scope of the current paper, as our focus is on the architectural presentation of the proposed service.

4. Semantics

Semantics contribute to the thematic organization of information on specific scientific fields. This is achieved by establishing a descriptive and controlled interrelated terminology. The establishment of structured and distinct terms of reference provides for the exchange of data in machine-readable form and therefore for interoperability. Moreover, semantics ensures our efficient access to structured information. Multiple indexes are being developed to secure high-quality search and retrieval. The development of such indexes[28.] presupposes the proper structuring of semantics according to international standards and formats.

Describing an item thematically is considered to be a subjective procedure, as the given information is not within the described item. Therefore, due to "semantic heterogeneities, resulting from different terminologies and conceptualizations,"[29. and 30.] the so-desired

(and necessary) compatibility seems to be almost an unaccomplished objective, as data-sharing bases on such schemas. [31.] This problem may be tackled through the development of knowledge management tools (i.e., thesauri, ontologies, knowledge bases) that [30., 31. and 32.]

• Establish descriptive terminology; a controlled vocabulary ensures the usage of common terms of reference, contributes to thematic description and hence to semantic compatibility.
• Guide the end-user to an efficient pathway towards high-quality information retrieval by depicting the established thematic terms and representing the hierarchical relationships between concepts.
• Organize scientific information and knowledge. The development of scientific fields has produced a vast new terminology, necessary for the management of the information on such fields. Thesauri play a key role as they use the bibliography of specific fields and provide the relevant terminology. In fact, subject terms need to be derived from relevant scientific items that are described in the bibliographic database. After all, this has been the practice of indexers for many years. It provides accuracy in vocabulary; it respects the scientific approach; it reflects the currents and structure of specific scientific fields; and it helps to ensure the currency of, and the continued evolution of, terminology. Building, then, a thesaurus can undoubtedly contribute not only to a more efficient retrieval, but also to the actual organization and terminology representation of a scientific field.

Thesauri are controlled vocabulary of terms in natural language that are designed for post-coordination.[33.] Taxonomies are a central part of a conceptual model and bring substantial order to elements of that model, present limited views for human interpretation, and play a critical role in reuse and integration tasks. [34.] Although both semantic schemes are used to structure semantic information in a domain, such knowledge organization tools lack semantic interoperability among different information systems. [35.] However, they do ensure data homogeneity within the same system. It is possible to integrate different thesauri and taxonomies and thus integrate information from different sources through ontologies. [35.] An ontology can be defined as "a logical theory accounting for the intended meaning of a formal vocabulary, i.e., its ontological commitment to a particular conceptualization of the world. The intended models of a logical language using such a vocabulary are constrained by its ontological commitment. An ontology indirectly reflects this commitment (and the underlying conceptualization) by approximating these intended models." [36.] Ontologies are characterized by their ability to express the conceptualization of the terms used to name a term of reference or thematically describe an item. [37.] Ontologies provide for common vocabulary to facilitate information-sharing and exchange. [38.] Hence, ontologies tend to play a major role in knowledge management, as they are able to represent not only the diction of the terms, but also their meaning and concept. [12. and 31.] Ontologies are the medium for making semantics both understandable by humans and machines. Hence, ontologies can ensure that the goal of semantics is fulfilled.

4.1. Encoding the semantics of government information

The index terms derived from the digitized documents constitute an authority file, while their thematic interlinking forms a multidimensional semantic network. The nodes of the semantic network are the index terms and they are interlinked using a thesaurus standard. In particular, two types of term associations have been realized: thematic and geospatial. The thematic associations link topical terms that are thematically and administratively related. The geographic associations provide for places of the state that conduct some form of

administrative activity and are interlinked with public services (corporate bodies that reside in the area). According to Fig. 2, the names of public servants that constitute the personnel of governments are associated with the name of the public service to which they are affiliated, the post that they hold, and their rank in the personnel hierarchy. The topical terms refer to posts, ranks, and competencies. Posts are connected with public services, names of public servants, and ranks. Ranks are linked with names of public servants and posts. Competencies are interlinked with names of public authorities. Finally, corporate bodies are all authorities of the public sector, as already mentioned, and are associated with geographic names (of areas in which they reside), posts, names of public servants that constitute the personnel of the organization/institution, and competencies of the public corporate body.

These relationship indicators, provided by the thesauri standard, are ideal for encoding PSI in its entirety, as they thematically describe the government publications by establishing descriptive terminology and they preserve the thematic and geospatial hierarchy. Thesauri cover the semantic relationships of equivalence (synonymy) by introducing the relationship indicators of "use" and "used for," of hierarchy by introducing the relationship indicators of "broader term" and "narrower term," and finally of association by introducing the relationship indicator of "related term."

However, it is essential to extract from the semantic network the hierarchy in the public sector, along with its competencies. This detachment results in a taxonomy schema that (1) facilitates information retrieval by providing two separate knowledge management tools, one for the depiction of a consensual representation of government information and the other for the mapping of administrative authorities and functions; and (2) information search, as the user may be guided and/or navigated within the public sector hierarchy. Moreover, the user becomes more familiar with the knowledge structure and how it is built within the system. Browsing for information may broaden or limit the submitted query, as the information provided is explicitly represented in the hierarchical structure.

4.2. Representing the public administration hierarchy

The public sector calls for term interlinking to represent geographic, thematic, and administrative hierarchies based on administrative functions.[39.] The first two types of hierarchy, geographic and thematic, may be represented via thesaurus relationship indicators, while the latter results from the semantic network and forms a taxonomy—a semantic schema capable of expressing multiple hierarchical levels. In the semantic network, the term used to represent a post, a competence, or a corporate body may be represented and associated thematically/conceptually but is not assigned to the required level of the administrative hierarchy to which it refers. For example, in a query for the term "Librar*" (with right truncation, library or libraries), the thesaurus will respond with the entry presented in Fig. 3.

Instead, in the same query, the taxonomy will first provide the number of organizations that are, or hold, a library (e.g., fifty organizations). Query limitations, such as the supervising authority of the libraries located, will secure accurate and immediate retrieval of the requested results (e.g., twenty libraries are supervised by the board of the parent organization and thirty libraries are supervised by the Special Secretariat of Libraries, under the Ministry

of National Education and Religious Affairs). Based on query settings, the taxonomy will provide the whole organizational chart that indicates the hierarchical scale of each retrieved library, along with its competencies and supervising/corporate authority. Table 1 presents a snapshot of the required information for the query "Librar*" submitted to the taxonomy. The system will retrieve either the library(ies) that constitute(s) an organization (e.g., National Library) or the library(ies) that administratively belong(s) to another corporate body (e.g., Library of Museum of Classical Arts). Moreover, the taxonomy will provide the subclasses of the retrieved entities along with their competencies and supervising authorities.

Hence, administrative hierarchy can be attained by creating a taxonomy from the semantic network while the thesaurus provides for conceptual interlinking. However, it is worth mentioning that with thesauri, it is difficult to map the organizational chart of authorities (i.e., multiple hierarchical levels) along with their functions and corporate/supervising organizations due to a lack of indicators. For example, using a "related term" indicator for administrative functions may result in a system conflict and/or confuse the user, as this indicator is also used for stating the terms that are thematically/conceptually connected. Alternatively taxonomies, via properties and constraints, may indicate the nominated competencies and supervising organizations.

Hence, the proposed semantic network and taxonomy consist of an ontology that facilitates knowledge management and provides navigation tools. Furthermore, the ontology ensures semantic interoperability among different information systems, whether they use the same vocabulary or not, since the ontology provides for a common conceptualization of that vocabulary.[36.] Within this framework, the ontology has been developed from the knowledge base following a bottom-up approach, providing "the means for describing explicitly the conceptualization behind the knowledge represented in a knowledge base." [40.] According to the data flow diagram of Fig. 4, the ontology produces results by parsing and filtering the semantic network nodes and edges. The ontology is created in the following way:

a. The algorithm searches the semantic network for nodes indicating corporate body names, entered either under place or jurisdiction (e.g., Athens. Internal Revenue Service, Kalamata. Internal Revenue Service, etc.) or under proper name (i.e., Ministry of Economics, National Library of Greece).
b. The detached nodes are further organized into a hierarchical structure according to their "vertical" associations (broader and narrower terms).
c. Then the algorithm searches the semantic network for the competencies that are further connected with the authority(ies) responsible.

5. Implementation

The Department of Archive and Library Sciences at the Ionian University, in cooperation with the National Documentation Centre of Greece, has undertaken the task of encoding and describing the first type of government information—legislation—which is currently held on microfilm. This scientific group consists of both information and administrative scientists.

The latter provide analysis of research results and government information regarding administrative procedures or terms of reference to aid information scientists in developing the semantic network. Programmers and technical assistants contribute to the implementation of the proposed model and resolve any technical problems arising.

The Unimarc format has been selected and implemented as the standard for the encoding and description of the Official Gazette's bibliographic data, and the Unimarc Authorities format for the creation and function of authority files (see Appendix A for the structure of metadata). [41.] Each legal act consists of a descriptive bibliographic record, which is linked to the record of the entire official gazette and the record of the microfilm to which it belongs. Each legal act has further information on its title and statement of responsibility, the publication date, the language(s) of text, and physical data (i.e., number of pages, tables, etc.). Until now, 5,943 bibliographic records and 16,500 authority records have been created over a ten-year period.

The thematic terms constitute an authority file in which every topical term, geospatial term, personal name, or corporate body constitutes a record. The interlinking of all terms within the authority file is based on thesaurus principles and relations. Therefore, we have integrated the metadata base with the semantic network by embedding the thesaurus associations within the authority file. This integration is obtained by using specific fields of the Unimarc Authorities standard to implement the thesaurus interlinks. In this way, the thesaurus functions within the same software environment as the metadata base (see Appendix B for the authority file and thesauri integration). This integration is significant because it ensures accuracy in thematic description and efficient enrichment of the semantic network. Thus, the update of the authority file permits the coherent enlargement of the semantic network. Hence, by applying the thematic grouping of terms for item description not only to retrieval but also the cataloguing phase, a better conceptual representation and retrieval is achieved.

Although previous work[42. and 43.] has addressed the need for semantic representation and management within the metadata environment, most of the known databases only provide thesauri (e.g., the ERIC database) or ontologies (e.g., Gaz-Guide [44.]) for retrieval purposes without being an integral part of the database due to different implementation standards. The standards used for item and collection description, such as Dublin Core, EAD, and MARC, make no provision for the development of a semantic schema within such a system. For example, the field used in Dublin Core for thematic description recommends the selection of a value from a controlled vocabulary or formal classification scheme but makes no provision for the structure of a thesaurus or ontology within the same environment. Dublin Core may have been used for metadata production within digital libraries but it was primarily designed for the description of networked resources. [26. and 27.] Within the same framework, EAD provides for the "designation of MARC encoding analogs and authorized form" via the handling of attributes but sets no guidelines for the development of an authority file. Moreover, another attribute defines the source of the controlled vocabulary terms used, as does Dublin Core. Finally, an element (called <controlaccess>) can enable authority-controlled searching across finding aids [45.] on a computer network, as well as the replication of collection-level search terms found in the 1xx, [46.] 6xx, [47.] and 7xx [48.] fields of MARC catalogue records in a finding aid. Finally MARC, and especially Unimarc format, contains a control subfield which, depending on the given value, may represent broader or narrower terms as required. This control subfield is a one-character alphabetic code that indicates a specific relationship between a tracing and a heading. However, this subfield is used in conjunction with reference tracing blocks. Therefore, the meaning of the

broader and narrower terms used in Unimarc is slightly different from that prescribed by thesaurus standards.

For the depiction and management of the taxonomic structure of our ontology, we have adopted the software provided by the Protégé Project.[38.] This software is a user-friendly tool, although its knowledge engineering presupposes inflexible management of the ontology entities. The following characteristics may be defined:

• The administrative hierarchy must be represented in classes and subclasses, using the relationship indicators "is a" and "part of." For example, Fig. 5 shows the main classes and subclasses of public administration.
• Valid instances for the above classes and subclasses could be any of the public services/authorities of the public sector that constitute a corporate body. For example, an instance of the class "ministries" might be the "Ministry of National Education and Religious Affairs" or the "Ministry of Economics." The National Library of Greece, mentioned in the first example, is considered to be an instance of the subclass "Legal Entities under Public Law."
• For each corporate body, which is an instance of a class or subclass, two types of slots[49.] are identified: competence and corporate/supervisory public authority. The competence slot describes the function and defines the identity of a public authority. This slot has a "string" facet and takes the value of the name of a competence(s). The cardinality [50.] of this slot is determined to be at least one. The corporate/supervisory public authority slot with value type "instance" also has a minimum cardinality of one. According to ontology principles, [37.] properties of a class, expressed via slots, might be instances of other classes within the same ontology. In other words, ontologies allow the definition of relationships between individuals by a facet called "instance-type." In such a case, the values of this slot are derived from the instances of other classes. In our case, the names of the corporate bodies that collaborate with the described identity or function as supervisory authorities are instances of the ontology classes and properties of the described public authority.

For example, the Special Secretariat of Libraries of the Ministry of Education that supervises the National Library is a slot of the class "National Library." The facet takes the value "instance-type" as the slot's values are instances of another class (i.e., the Special Secretariat of Libraries is an instance of the subclass "administrative units" of the class "ministries").

It is essential for a competence slot to be allocated to the specific level of the administrative hierarchy that is responsible for the accomplishment of the given function(s) due to a hereditary characteristic that emphasizes the hierarchical nature of ontologies (all subclasses of a class inherit the slot of that class). This parameter further demonstrates why thesauri lack the ability to represent administrative hierarchies and organizational charts.

Fig. 6 indicates how the administrative hierarchy of a public authority may be represented via ontology principles and which slot types are required to describe its properties. According to our procedure, the creation of the ontology is as follows:

a. The algorithm searches the semantic network for nodes indicating corporate body names entered either under place or jurisdiction (e.g., Athens. Internal Revenue Service, Kalamata. Internal Revenue Service, etc.) or under their proper name (e.g., Ministry of Economics,

National Library of Greece). This search takes place in field 210 ("corporate body name") of the authority file. The algorithm searches only for the field 210 and assigns the first indicator[51.] the value 0—corporate name—and the second indicator, [52.] with either the value 1, if the name was entered as a place or jurisdiction, or 2, if the name was entered as its proper name.
b. The detached nodes are further organized into a hierarchical structure according to their "vertical" associations (broader and narrower terms). The algorithm detaches the 210 fields along with their broader and narrower links and builds a taxonomy hierarchy by developing a "component/part of" relation between the 210 fields and their vertical connections.
c. Then the algorithm searches only for competencies, which become a slot. The facet for this slot is determined to be "string" and its cardinality takes the value "multiple." The algorithm searches only for those under field 250 (topical subject) of the authority file fields that are linked with the extracted 210 fields. The system detaches only those competencies that are connected with those public services represented in the ontology hierarchy. The extracted competencies with field 250 become a slot. The facet is determined to be a "string" and the cardinality is at least one.

The corporate or supervisory public authorities of the public services, already established as classes or subclasses in the ontology hierarchy, are considered to be slots. The facet is defined to take the type value "instance-type,"[53.] as the slot's values are instances of another class, and the cardinality is determined to be at least one.

The representation of the administrative hierarchy via an ontology is currently under development. As already mentioned, the ontology is created by parsing the data of the authority file and detaching those terms that correspond to the administrative hierarchy of the public sector. "Generating ontological metadata from MARC is a form of preprocessing, in which relationships implicit in MARC are converted into explicit, labeled relations amenable to manipulation by computers."[54.] Therefore, ontology is used to extract the terms describing public authorities and to organize them in a manner that reflects the hierarchy of the public sector.

6. Related work

Government information needs immediate encoding of administrative material, as the high quality of record management ensures efficient access to required information both for government employees and citizens. There are many ongoing initiatives that are currently focused on methods and principles of providing for better description and encoding of the data produced by the public sector.

Many countries are pursuing the development and adoption of a common policy for record management, as they aim for electronic governance by 2004. Within this framework, they have started examining several platforms that will provide a stable environment for the efficient management of electronic records. Such initiatives are those currently running in Great Britain, New Zealand, Australia, and Canada. The technical characteristics of these initiatives are mentioned in the Syntax issues and description standards section. Briefly, we shall mention that these systems are mainly concerned with public records and their

transaction within central, regional, and local administration and not with the description and encoding of all government information.

Within New Zealand's e-government program, a specialized project named the Portal Thesaurus Project[55.] aims to develop and select thesauri to describe government services and information resources. Two separate, built-up, thesauri are being created; the first for representing functions and the second for establishing subject terms. In our system, the proposed thesauri are used to represent geographic and thematic hierarchies as well as to establish descriptive terminology, derived from the digitized texts by implementing a bottom-up procedure. The administrative hierarchy, along with the functions of public authorities, is represented via an ontology schema, which is produced by automatically parsing the semantic network and extracting the required semantics.

Many researchers as well as projects have been oriented towards providing an efficient architecture for the description and encoding of government information. The Gaz-Guide project[44.] proposes a multiagent system that assists information retrieval of the official gazette by mediating for the user's information needs and the semantic structure of the data domain. This system embeds both the ontology and the thesaurus to traverse different cognitive spaces. In the Gaz-Guide platform, the proposed thesaurus refers only to topical subjects, while the ontology represents (a) the types of legal acts as well as any kind of revision or amendment and (b) data on the issuing authority of each legal act (including contact information, URLs, functions, and hierarchical relations). Our architecture refers to the description and encoding of all available government information (legislation, the official gazette, and public records) and focuses both on the data/metadata structure and on information retrieval. Specifically, we deal with the representation of the geographic and thematic hierarchy, along with the established descriptive terminology on the scientific domain of the public sector, and the public sector hierarchy and its functions. Moreover, in our architecture, all terms in the thesauri are interlinked—based on administrative functions—and operate within an authority file that resides in the same environment in which metadata are structured and maintained. Furthermore, the type of the legal act is represented in the bibliographic metatags, as such information refers to the bibliographic content of documents and not to semantics. Finally, we suggest how an ontology may be developed and structured in a taxonomic schema by parsing thesauri indicators and extracting the required identities and concepts, which are continuously being developed and reorganized in the hierarchy of the public sector.

Within the same framework is another project that proposes a software platform that provides access to administrative information resources.[6.] The application domain concerns the scientific field of health. The project focuses on how end-users may ensure access to information sources on the Internet, as well as on how the retrieved information may be represented. An ontology is used to describe concepts to facilitate information retrieval. The main difference with our ontology is that via a taxonomic schema, it represents the administrative hierarchy and not the thematic hierarchy which, in our model, is organized in a thesaurus-like semantic network. Moreover, our focus is on the primitive structure of data and metadata and not on tracing and organizing information on the Internet.

Similarly the EUSlanD[56.] project suggests a methodology for organizing and classifying government information by developing application packages for information exchange in local and regional government. These packages are based on tagged abstracts for bibliographic description and on thesaurus schema for semantic representation.

The Energy Data Collection (EDC) system[57.] supports homogeneous access to multiple energy databases. The EDC system uses ontology to unify the metadata and domain terms of different databases and therefore to create a coherent domain model. Specific domain ontologies are integrated and mapped in a higher ontology. Therefore, ontology principles are only used to taxonomize sets of terms derived from different sources.

There is yet another approach in which a given initial legal ontology is refined using an existing machine-readable dictionary.[9.] Similarly, lexical knowledge bases may be developed from machine-readable dictionaries using ontology principles and methods. [58.] However, these approaches are partial attempts that focus only on establishing legal terminology using ontologies and machine-readable dictionaries.

Concerning developments in e-government, the Smart-Gov project[8.] provides a knowledge-based platform for transaction services via e-forms. However, the project has not yet defined system specifications for each component of the knowledge base. Here, we propose a semantic global network for all government information, and we determine the principles for an appropriate data and metadata structure, which ensure efficient and high-quality retrieval for any on-sight service. Therefore, our approach may also be used as a basis for the development of transaction services.

Finally other approaches, such as the KIWI,[7.] On-to-Knowledge, [59.] and OntoWeb [10.] projects, as well as Storey et al., [11.] use ontologies to present semantics and facilitate knowledge management either within organizations or the Web. However, these initiatives do not focus on complete platforms for the organization and delivery of government information in its entirety.

7. Conclusions

The increasing demand for government information calls for advanced knowledge-based framework systems for information gathering, flow, and distribution. Due to the public sector's hierarchical nature, government information, produced or disseminated by central government authorities, is fragmented and dispersed. Current initiatives do not meet the growing information needs, as they lack the advanced information management tools necessary to secure data and metadata homogeneity, as well as high-quality information search and retrieval. The proposed digital architecture, based on international standards and formats for its structure, semantics, and syntax, provides for efficient and accurate retrieval via the function of multiple indexes that permit simultaneous queries in different fields, data exchange, and metadata homogeneity, as common standards and formats have been adopted, and proper flow and dissemination of government information through administrative channels of central, local, and regional government hierarchy, as the implemented knowledge management tools depict, represent, and encode the administrative hierarchy of the public sector and facilitate user navigation through structured information. The depiction of an administrative hierarchy provides for better public services, as citizens may obtain access to the required information promptly and precisely; advanced insight services, as public servants execute their transactions with accuracy and efficiency; and high-quality services for policy and decision makers, as the system presents an integrated depiction of the public sector both in hierarchy and function. Moreover, the proposed digital architecture may constitute the basis for the development and function of government portals and e-government applications due to the possibility of one-stop information rendering.

The integration of thesauri within authority files and the usage of terms instead of subject headings ensure thematic homogeneity and accuracy, as this leads to thematic allocation and therefore aids the end-users in their searches. Furthermore, all thematic terms are linked, based on administrative and geospatial connections that are indicated by administrative functions and not solely by thematic association and relevance. Therefore, the thesauri interlink heterogeneous terms such as personal names, corporate bodies, and geospatial or topical terms. Moreover, the developed geographic chain within an authority file indicates how public services are disseminated and also identifies the local/regional administrative authority they come under. The integration of thesauri within authority files contributes to the production of the semantic network within the authority file.

The developed taxonomy and ontology, derived from the semantic network, represent the administrative hierarchy of the public sector. The taxonomy contributes to the development of a classification scheme for government information by implementing an algorithm based on administrative hierarchy. Such classification aids the execution of specialized searches, the exchange of information and metadata, and the issuing of scientific printouts in certain domains.

The integration of thesauri and taxonomies depicts and represents all the available governmental knowledge and ensures accurate and quality metadata, as well as interoperability in the midst of legislative, administrative, and geospatial information diversity. Moreover, the ontology ensures semantic homogeneity by stating the conceptualization of the knowledge base.

The proposed knowledge management model is being developed and updated by implementing a bottom-up process. Hence, the digitized texts provide the terms/concepts that will be used for thematic description. These terms, organized by thesaurus principles within the semantic network, provide the entities (public services) and their hierarchy, which further provides the proposed taxonomy and hence the ontology. Due to the nature of the public sector, alterations and changes may be observed in the administrative hierarchy or in the competencies of services. The proposed knowledge management model meets this need due to its dynamic characteristic that permits immediate and secure updating. Since the proposed taxonomy results from the semantic network and is embedded within the authority file, they may locate any changes or additions to the structure and/or functions of the public sector.

Data homogeneity and system interoperability are considered to be essential for proper and efficient metadata exchange, semantics, or syntax. Within this framework, researchers are aiming to adopt common standards and formats to secure quality metadata. However, current initiatives refer to the expansion of existing standards and therefore contribute to content diversity. In the near future, we will examine how different syntax standards for information description and encoding may be integrated under the same platform and how this integration ensures quality metadata and a homogeneous syntax environment.

References

1. Government policy on archives. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.pro.gov.uk%252Farchives. Accessed January 15, 2003.

2. European Commission, Public sector information: A key resource for Europe. (1998) Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fmineco.fgov.be%252Finformation_society%252Fadministrations%252Fps_034_en.pdf. Accessed January 15, 2003 .

3. Australian Legal Information Institute. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.austlii.edu.au; CALI. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Flessons.cali.org%252Fcatalog.html; Cornell's Legal Information Institute. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.law.cornell.edu; Lexis Nexis. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.lexis.com%252Fresearch; Lois Law. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.loislawschool.com; National Printing-Office of Greece. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.et.gr; Tax Treaties Online. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.oceanalaw.com%252Fdefault.asp; UN Treaty Collection. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Funtreaty.un.org%252Fenglish%252Faccess.asp; West Law. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.lawschool.westlaw.com. Accessed January 15, 2003.

4. Find Law. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.findlaw.com; Guide to Online Law. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.loc.gov%252Flaw%252Fguide%252Findex.html; Internet Legal Resource Guide. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.ilrg.com;

Jurist. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fjurist.law.pitt.edu; Law.Com. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.law.com; LLRX. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.llrx.com. Accessed January 15, 2003.

5. T. Koch, Quality controlled subject gateways: Definitions, typologies, empirical overview. *Online Information Review* **24** (2000), pp. 24–34. Abstract-Compendex | $Order Document | **Full Text** via CrossRef

6. A. Anagnostakis, G. Sakellaris, M. Tzima, D. Fotiadis and A. Likas, Citation: Citizen information tool in smart administration. In: *Proceedings of the 1st International Conference on Electronic Government, EGOV 2002Lecture Notes in Computer Science, (LNCS 2456)*, Springer-Verlag, Berlin (2002), pp. 307–312. Abstract-INSPEC | $Order Document

7. L. Gadda, E.B. Innocenti and A. Savoldelli, KIWI: Building innovative knowledge management infrastructure within European Public Administration. In: *Proceedings of the 1st International Conference on Electronic Government, EGOV 2002Lecture Notes in Computer Science, (LNCS 2456)*, Springer-Verlag, Berlin (2002), pp. 223–229.

8. P. Georgiadis, G. Lepouras, K. Vassilakis, G. Boukis, E. Tambouris, S. Gorilas, E. Davenport, A. Macintosh, J. Fraser and D. Lochhead, Smart Gov: A knowledge-based platform for transactional electronic services. In: *Proceedings of the 1st International Conference on Electronic Government, EGOV 2002Lecture Notes in Computer Science, (LNCS 2456)*, Springer-Verlag, Berlin (2002), pp. 362–369. Abstract-INSPEC | $Order Document

9. M. Kurematsu and T. Yamaguchi, A legal ontology refinement support environment using a machine-readable dictionary. *Artificial Intelligence and Law* **5** (1997), pp. 119–137. Abstract-Compendex | Abstract-INSPEC | $Order Document | **Full Text** via CrossRef

10. Ontoweb: Ontology based information exchange for knowledge management and electronic commerce. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fontoweb.aifb.uni-karlsruhe.de. Accessed January 15, 2003.

11. V.C. Storey, D. Dey, H. Ullrich and S. Sundaresan, An ontology-based expert system for database design. *Data and Knowledge Engineering* **28** (1998), pp. 31–46. SummaryPlus | **Full Text + Links** | PDF (903 K)

12. Content Standard for Digital Geospatial Metadata. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.fgdc.gov%252Fmetadata%252Fcontstan.html. Accessed January 15, 2003.

13. Y. Ding, D. Fensel, M. Klein and B. Omelayenko, The Semantic Web: Yet another hip?. *Data and Knowledge Engineering* **41** (2002), pp. 205–227. SummaryPlus | **Full Text + Links** | PDF (407 K)

14. Each post within the public sector (held by a civil servant) corresponds to a rank and a position classification, which actually represents the specialization of each employee.

15. UNIMARC is a format used to facilitate the international exchange of bibliographic data in machine-readable form between national bibliographic agencies. UNIMARC may also be used as a model for the development of new machine-readable bibliographic formats. UNIMARC is maintained by an IFLA committee, the Permanent Unimarc Committee (PUC), for which the secretariat is the IFLA UBCIM Core Programme; UNIMARC Manual. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.ifla.org%252FVI%252F3%252Fp1996-1%252Fsec-uni.htm. Accessed January 15, 2003.

16. An analytic table of all the fields used in the encoding of legislation is provided in Appendix A.

17. *ISO 15489: Records Management—General*. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.naa.gov.au%252Frecordkeeping%252Frkpubs%252Ffora%252F02Mar%252FAS_ISO_15489.pdf. Accessed February 6, 2004.
S. Healy, ISO 15489 records management: its development and significance. *Records Management Journal* **11** 3 (2001), pp. 133–142. **Full Text** via CrossRef

18. The Dublin Core Metadata Initiative is an open forum engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models. Further information may be found at http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fdublincore.org%252F.

19. The EAD Document Type Definition (DTD) is a standard for encoding archival finding aids using the Standard Generalized Markup Language (SGML). The standard is maintained in the Network Development and MARC Standards Office of the Library of Congress (LC) in partnership with the Society of American Archivists. Further information may be found at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Flcweb.loc.gov%252Fead%252F; Dublin Core Metadata. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fdublincore.org%252F. Accessed January 15, 2003; EAD (Encoded Archival Description) Document Type Definition. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Flcweb.loc.gov%252Fead%252F. Accessed January 15, 2003.

20. GILS—Government Information Locator Service. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.gils.net. Accessed January 15, 2003.

21. E-GMS (UK Government Metadata Standard). Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.govtalk.gov.uk%252Fschemasstandards%252Fmetadata.asp. Accessed January 15, 2003.

22. E-GIF (UK Government Interoperability Framework). Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.govtalk.gov.uk%252Fschemasstandards%252Fegif_document.asp%253Fdocnum%253D731. Accessed January 15, 2003.

23. GLS Metadata Standard (Australian Government Locator Service). Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.naa.gov.au%252Fdefault.html. Accessed December 20, 2002.

24. NZGLS Metadata Standard (New Zealand Government Locator Service). Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.e-government.govt.nz. Accessed January 15, 2003.

25. American National Standards Institute. (1975). *Representations of universal time, local time differentials, and United States time zone reference for information interchange (ANSI X3.51-1975)*. New York: American National Standards Institute; American National Standards Institute. (1975). *Representation for calendar date and ordinal date for information interchange (ANSI X3.30-1985)*. New York: American National Standards Institute; American National Standards Institute. (1975). *Representation of local time of day for information interchanges (ANSI X3.43-1986)*. New York: American National Standards Institute; American National Standards Institute. (1975). *Representation of geographic point locations for information interchanges (ANSI X3.61-1986)*. New York: American National Standards Institute; American National Standards Institute. (1975). *Dictionary for information systems (ANSI X3.172-1990)*. New York: American National Standards Institute.

26. O'Kane, T. (2001). *Mapping the similarities and differences between metadata standards: Dublin Core, AGLS and NZGLS: The GUIDE Project*. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.e-government.govt.nz%252Fdocs%252Fmetadata-standards-differ%252Fmapping-diff.pdf. Accessed February 6, 2004.

27. See http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.loc.gov%252Fmarc%252Fmarcxml.html.

28. High-quality indexes are considered those that consist of enhanced features ensuring not only a proper data structure, but also the requirements needed for easy update and maintenance.

29. B. Amann and I. Fundulaki, Integrating ontologies and thesauri to built RDF Schemas. In: *Proceedings of the 3rd European Conference for Digital Libraries ECDL'99Lecture Notes in Computer Science, (LNCS 1696)*, Springer-Verlag, Berlin (1999), pp. 234–253. Abstract-INSPEC | $Order Document

30. A. Sheth, Changing focus on interoperability in information systems: From system, syntax, structure to semantics. In: M.F. Goodchild, M.J. Egenhofer, R. Fegeas and C.A. Koffman, Editors, *Interoperating geographic information systems*, Kluwer Academic Publishers, Dordrecht (1999), pp. 5–29.

31. H. Pundt and Y. Bishr, Domain ontologies for data sharing: An example from environmental monitoring using field GIS. *Computers and Geosciences* **28** (2002), pp. 95–102. SummaryPlus | **Full Text + Links** | PDF (170 K)

32. T. Brasethvik and J. Gulla, A conceptual modelling approach to semantic document retrieval. In: *Proceedings 14th International Conference on Advanced Information Systems Engineering, CAISE 2002Lecture Notes in Computer Science, (LNCS 2348)*, Springer-Verlag, Berlin (2002), pp. 167–182.

33. International Organization for Standardization, Documentation-Guidelines for the establishment and development of monolingual thesauri (ISO 2788).

34. C. Welty and N. Guarino, Supporting ontological analysis of taxonomic relationships. *Data and Knowledge Engineering* **39** (2001), pp. 51–74. SummaryPlus | **Full Text + Links** | PDF (739 K)

35. M. Pisanelli, A. Gangemi and G. Steve, Ontologies and information systems: The marriage of the century?. In: *Proceedings of the Lyee workshop, Paris* (2002) Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.loa-cnr.it%252FPapers%252Flyee.pdf. Accessed February 6, 2004 .

36. N. Guarino, Formal ontology in information systems. In: *Proceedings of the International Conference on Formal Ontology in Information Systems, FOIS '98*, IOS Press, Amsterdam (1998), pp. 3–15. Abstract-INSPEC | $Order Document

37. B. Chandrasekaran, J.R. Josephson and R. Benjamins, What are ontologies and why do we need them?. *IEEE Intelligent Systems* **14** 1 (1999), pp. 20–26. Abstract-Compendex | Abstract-INSPEC | $Order Document | **Full Text** via CrossRef

38. Noy, N. F., & McGuiness, D. L. (2001). *Ontology development 101: A guide to creating your first ontology*. Stanford Knowledge Systems Laboratory Technical Report KSL 01-05 and Stanford Medical Informatics Technical Report SMI 2001-0880.

39. M. Whiteman, In the shadow of hierarchy: Meta-governance, policy reform and urban regeneration in the West Midlands. *Area* **35** 1 (2003), pp. 6–14.

40. A. Bernaras, I. Laresgoiti and J. Corera, Building and reusing ontologies for electrical network applications. *Proceedings of the European Conference on Artificial Intelligence, ECAI* **96** (1996), pp. 298–302. Abstract-INSPEC | $Order Document

41. D. Kyriaki-Manessi, D. Moschopoulos and G. Prokopiadou, Government documents information management: Applications to the official gazette. In: (2002), pp. 42–54.

42. Pollitt, S. (1997). *The key role of classification and indexing in view-based searching*. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.ifla.org%252FIV%252Fifla63%252F63polst.pdf. Accessed January 15, 2003.

43. Taylor, M. *Zthes: A Z39.50 profile for thesaurus navigation*. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.lcweb.loc.gov%252Fz3950%252Fagency%252Fprofiles%252Fzthes-04.html. Accessed January 15, 2003.

44. J.S. Liu, V.W. Soo, C.N. Ching, C.Y. Lee and C.Y. Lin, Gaz-guide: Agent-mediated information retrieval for official gazettes. In: *Proceedings of the 4th Pacific Rim International Workshop on Multi-Agents, PRIMA, 2001Lecture Notes on Artificial Intelligence, (LNAI 2132)*, Springer-Verlag, Berlin (2001), pp. 154–167.

45. Finding aid is the broadest term to cover any description or means of reference made or received by an archives service while establishing administrative or intellectual control over archival material (The definition is derived from the ISAD (G): General International Standard Archival Description standard).

46. This block is named Main Entries and contains a name or a uniform title heading, which is used as main entry.

47. This block is named Subject Access Fields and contains subject access entries and terms.

48. This block is named Added Entries and provides additional access to a bibliographic record from names and/or titles having various relationships to an item.

49. Slots define properties of classes.

50. The slot cardinality defines how many values a slot can have.

51. This indicator is used to specify the kind of corporate body (i.e., corporate name or meeting).

52. This indicator is used to specify the way the name of the corporate body is entered.

53. This value-type allows definition of relationships between individuals. The slots' values are instances of classes.

54. P.C. Weinstein and W.P. Birmingham, Creating ontological metadata for digital library content and services. *International Journal on Digital Libraries* **2** (1998), pp. 20–37. Abstract-INSPEC | $Order Document | **Full Text** via CrossRef

55. Interim report of the portal thesaurus project. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.e-government.govt.nz%252Fdocs%252Finterim-thesaurus%252F. Accessed January 15, 2003.

56. EUSlanD Project. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.eusland.org%252Fhtml%252Fw5.html. Accessed January 15, 2003.

57. E. Hovy, Using an ontology to simplify data access. *Communications of the ACM* **46** 1 (2003), pp. 47–49. Abstract-Compendex | Abstract-INSPEC | $Order Document

58. B. Whiteman and J. Klavans, Extracting taxonomic relationships from on-line definitional sources using LEXING. In: *Proceedings of the 1st ACM-IEEE CS Joint Conference on Digital Libraries*, ACM Press, New York (2001), pp. 257–258.

59. On-to-Knowledge Project. Available at: http://www.sciencedirect.com/science?_ob=RedirectURL&_method=externObjLink&_locator=url&_cdi=6542&_plusSign=%2B&_targetURL=http%253A%252F%252Fwww.ontoknowledge.org.