Encoding of Public Records and Procedures in Electronic Registries

Georgia Prokopiadou, Christos Papatheodorou, Dionysis Moschopoulos

Department of Archive and Library Sciences, Ionian University
Palea Anaktora, 49100, Corfu, Hellas
{gproko, papatheodor, moschop}@ionio.gr

Abstract. An electronic registry is the main mechanism for the official distribution of public records, which are created within the framework of administrative procedures and business transactions. The encoding of public records within a registry is realized according to a certain format to facilitate information management. However, these formats do not refer to semantics, which are considered essential for: a) the description of the administrative procedures, and b) efficient information retrieval. This paper aims to provide a digital library architecture for the encoding of public records and administrative procedures. It introduces the usage of semantic tools for the management of knowledge within the proposed integrated system.

1. Problem Statement

The main condition for the development of government information systems is the ensuring of interoperability for public services to exchange data in machine-readable form. Hence, worldwide interoperability frameworks (e.g. NZ e-GIF [10], UK e-GIF [15]) and thesauri (e.g. EUROVOC [3], FONZ [4], SONZ [14]) have been developed, in order to determine the required information matrix for the encoding of public records. However, public records management requires the representation of administrative procedures and transactions, along with their semantics and relations.

Administrative transactions refer to dealings with professionals or citizens, as well as dealings that are realized exclusively between public services and are intended to meet the fundamental needs of public administration. The first category may be divided into: a) the transaction that a citizen/professional has with the responsible public service (G2C), given the fact that he/she needs to be informed about the documentation requested, and/or submit an application in order to initiate the procedure, and b) the internal procedures within the public service (G2G) independently of citizen/professional's request. G2C transactions maintain personal interest and therefore the search and retrieval of information is based on personal criteria. On the other side, G2G transactions require the existence of several access points since an information search needs to be based on the semantic correlations between records. However, both transactions are part of an information and workflow cycle that may be distinguished according to the following five stages that a public service needs to follow: a) accomplish the required tasks by creating the records

needed, b) handle the delivery of the file either to internal departments or to external services, c) manage and control the transactions, d) update the database regarding the service that accesses / creates the file in a given instance, and e) carry out any transactions necessary for the completion of the administrative action.

Currently, public services use different systems for record-keeping and workflow supervision. Thus, electronic registry meets only the first two stages of the information life cycle, while the remaining stages are managed and controlled by specialized workflow systems.

Our proposal ensures the management of public records and procedures within the same information context enforced by semantic indexes. In particular the main objective of our approach is to: a) modulate an advanced registry for the monitoring of public records using thesauri, taxonomies and ontologies, and b) to encode the administrative procedures, as well as the workflow of public services within the same registry as public records. The proposed architecture introduces the creation of taxonomies for the structure of public administration and uses these taxonomies for the monitoring of public records and administrative procedures.

2. Information Needs

2.1 Bibliographic Information Metadata

Bibliographic metadata are mainly used to uniquely identify a public record by encoding, for example, its title, the date of mail / delivery, the register number, as well as the name of sender(s) and receiver(s). They may be described implementing the ISO 15489 standard, which defines the technical specifications for the management of public records and classifies the criteria for their efficient control [7].

Bibliographic metadata encoding for public records may be realized by adopting and implementing standards such as Dublin Core, which is considered to be an open standard for the development of interoperable online metadata [2]. However, the offered meta-fields cannot describe government information in its entirety, as only a 15-element set of descriptors [9] are offered. Over and above indicative meta-fields for the encoding of the official records, there may be: a) the administrative "functions" of the Service from which the public records are derived, b) the "availability" of the records and how they can be obtained, c) the "audience", i.e. the target group of the encoded official records, d) the "mandate" that requires the record to be created, i.e. acts, regulations, rules, court cases, e) the conditions of their "disposal" and "preservation", and f) the "location" of the official records. Due to the lack of essential fields for the depiction of the aforementioned metadata, many countries have proceeded with an extension of the Dublin Core standard, in order to meet the information needs of public records. Such cases are: a) GILS (Government Information locator Service) [5] for United States of America, b) e-GMS (UK Government Metadata Standard) [16] for United Kingdom, c) AGLS Metadata Standard (Australian Government Locator Service) [1] for Australia, and d) NZGLS Metadata Standard (New Zealand Government Locator Service) [11] for New

Zealand. Due to different extensions of the Dublin Core standard, the aforementioned countries proceeded to issue crosswalks, in order to ensure interoperability and system compatibility [8].

The depiction of the process and the accomplishment stages of an administrative procedure are considered as an essential parameter in the management of administrative data. However, public administration procedures are totally dependent on the public records that justify them. Public Records become an entity as soon as they are inserted in the registry. Therefore, the meta-fields for the encoding of administrative procedures need to be developed in absolute correspondence with the electronic registry. In these meta-fields it is essential to encode: a) the responsible organization and its qualified department/office for this procedure, b) the stages of processing required, c) the services involved, d) the stage of the transaction in a given moment, e) the public service processing the file at a given instance, and f) time estimation for the closure of a particular file.

The encoding of such information may be implemented only on a semantic basis. In particular, the values of these meta-fields need to be derived from knowledge management tools that depict the hierarchy of public services and their corresponding procedures.

2.2 Semantic Information Metadata

The proposed architecture implements two knowledge management tools:

- A semantic network built according to thesaurus principles, in order to: a) establish administrative controlled and interlinked terminology, and b) represent the hierarchy of administrative procedures as they depict the stages required for the fulfillment of an administrative action.
- A taxonomy for the: a) encoding and depiction of the administrative hierarchy of the public sector, and b) navigation by end users of the structured information representing the arrangement and relationships of each hierarchical level.

The proposed system interlinks: a) names of corporate bodies - public services in central, regional and local administration, b) geospatial names - places of state that fulfill some form of administrative activity, and c) topical terms which are conceptually related. Geospatial relations link names of places with public services and organizations holding some form of administrative activity. These categories represent the geospatial and thematic hierarchy of the public sector and form a semantic network. Apart from these hierarchies, government information also calls for representation and depiction of the administrative hierarchy, which refers to terms that need to be linked due to administrative activity and is essential to connect public services with procedures. However, the thesaurus may represent the conceptual and geospatial connections but is unable to depict multiple hierarchical levels each of them related with the proportional procedures. In particular, the development of administrative hierarchy within conceptual relations results in the usage of the same indicators (Broader Term, Narrower Term, Relation Term) for different concepts, i.e. topical terms and corporate bodies.

Consequently, the administrative hierarchy may be represented in a taxonomy schema capable of establishing multiple hierarchical levels all connected with the

corresponding procedures and public authorities derived from the semantic network. An algorithm has been proposed to detach the corporate bodies from the semantic network and to organize them into a hierarchical structure based on their vertical associations (broader and narrower terms) [13]. The proposed algorithm may also search and detach the entries of the procedures of public services. Within the scope of the taxonomy, these procedures are linked with the responsible hierarchical entity. Since the taxonomy derives from the semantic network, clearly the two knowledge representation and management tools interact and complement each other, providing an integrated framework for the management of government information. The coexistence of the thesaurus and taxonomy forms an ontology, which enables the management of the semantic information derived from public records. Moreover, ontology provides for navigation tools allowing the map-reading of the public sector's hierarchy tree. Furthermore, the ontology ensures data interoperability and policy compliance between different information systems, even if the latter do not adopt the same terminology, due to the fact that it provides for a common realization of the conceptual world [6].

The two knowledge management tools support the depiction of the administrative procedures. Specifically, the semantic tools accommodate values for the following elements: the responsible organization and its qualified department / office for an administrative procedure, the services involved, and the public service processing a file at a given instance, which are derived from the taxonomy schema. On the other side, the stages required for an administrative procedure are derived from both the thesaurus and the taxonomy since the thesaurus represents the administrative functions and these are correlated with 'services' in the taxonomy.

3. Architecture

Each public service handles and manages the administrative actions that correspond to its transactions. Therefore, the public authorities are responsible for: a) creating documents in the registry and connecting them with the corresponding transactions, and b) updating the registry regarding the stage of an action's fulfillment. This activity may be enforced by governments' initiative to unify the format of all public record types. Electronic registry development and update is realized by each public service separately, according to the in-house policy for public record management and organization. Within this framework, a network of digital registries is generated, within which every public service creates and updates a local digital library of public records.

Each digital library consists of: a) the digitized data of public records, b) the bibliographic metadata that uniquely identify public records and procedures, c) a semantic network, and d) a taxonomy.

According to Figure 1, the end-user can make queries concerning the follow-up of G2C or G2G transactions and their related records that have been inserted in the electronic registry of a public service. A harvester handles the query by crawling within the digital libraries metadata. The bibliographic metadata identify the records

related to the query, while semantic metadata provide the harvester with interlinked and correlated information regarding the administrative structure and procedures.

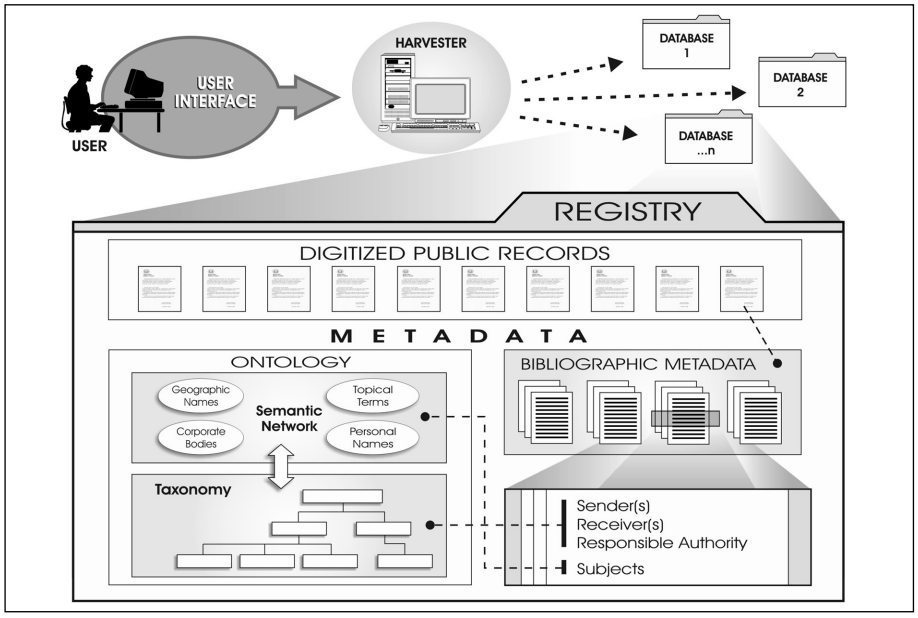

Fig. 1. The proposed architecture

By using harvesting protocol principles, the distributed electronic registries efficiently disseminate their content, as different repositories, while the harvester effectively collects the metadata residing in different digital libraries. According to Open Archive Initiative Protocol for Metadata Harvesting (OAI-PMH), Dublin Core is the common metadata format for interoperability on bibliographic metadata [12]. However, other metadata schemas (i.e. NZGLS, AGLS, etc.) can be harvested provided those are encoded in XML schema.

4. Conclusions

The main objective of this paper is to introduce a schema for the encoding of public records and procedures running within the same information environment, given the fact that the latter determine the creation and official distribution of public records. The proposed architecture of electronic registry for public record and transaction-encoding facilitates G2C and G2G dealings by providing advanced and upgraded services for information management and retrieval, as well as semantic interoperability. The most important features are the ability to: a) support decision-making procedures by registering all the required data (history of a service's actions, responsible and involved public authorities, etc.), and b) facilitate citizen-state and inner service administrative procedures by registering all the incoming and outgoing records, as well as a service's transactions. The development of a decentralized

network of electronic registries with a digital library structure possesses the following features:

- ensures multiple access points for the execution of sophisticated queries
- provides for user navigation through structured information
- supports the qualitative data entry and retrieval of public services, contributing to the better depiction and search of their hierarchical structure
- enables the location of data that reside in different information environments
- facilitates the control of the public administration's transactions
- holds out the development of a classification scheme, in order to execute specialized queries and printouts for the support of decision-making procedures
- contributes to data homogeneity and system interoperability, and
- adopts and maintains a common policy for record keeping, management and distribution.

References

1. Australian Government Locator Service [online]. Available at: http://www.naa.gov.au/recordkeeping/gov_online/agls/metadata_element_set.html
2. Dublin Core [online]. Available at: http://dublincore.org/
3. EUROVOC Thesaurus. [online]. Available at: http://www.europa.eu.int
4. Functions of New Zealand Thesaurus (FONZ) [online]. Available at: www.e-government.govt.nz/nzgls/ thesauri/fonz/fonz220.pdf
5. Government Information Locator Service [online]. Available at: http://www.gils.net
6. Guarino, N.: Formal ontology and information systems. In: Proceedings of FOIS ´98. IOS Press Amsterdam (1998) 3-15
7. Healy, Susan: ISO 15489 Records Management: its development and significance [online] (2001). Available at: http://mustafa.emeraldinsight.com/pdfs/rmj1.pdf
8. Kane, T. O': Mapping the similarities and differences between metadata standards: Dublin Core, AGLS and NZGLS: the GUIDE Project [online] (2001). Available at: http://www.e-government.govt.nz/docs/metadata-standards-differ/mapping-diff.pdf
9. National Information Standards Organization: The Dublin Core Metadata Element Set. NISO Press Maryland (c2001)
10. New Zealand e-Government Information Framework [online]. Available at: http://www.e-government.govt.nz/docs/e-gif-v-2/
11. New Zealand Government Locator Service [online]. Available at: http://www.e-government.govt.nz
12. Open Archives Initiative Protocol for Metadata Harvesting [online]. Available at: http://www.openarchives.org/OAI/openarchivesprotocol.html
13. Prokopiadou, G., Papatheodorou, C., Moschopoulos, D.: Integrating Knowledge Management Tools for Government Information. In Government Information Quarterly, Spring Issue (under publication). Elsevier Amsterdam (2004)
14. Subjects of New Zealand Thesaurus (SONZ) [online]. Available at: www.e-government.govt.nz/nzgls/ thesauri/sonz/sonz120.pdf
15. UK e- Government Information Framework [online]. Available at: http://www.nhsia.nhs.uk/egif/pages/default.asp
16. UK Government Metadata Standard [online]. Available at: http://www.govtalk.gov.uk/schemasstandards/metadata_document.asp?docnum=840