

DATA MODEL FOR STORAGE AND RETRIEVAL OF LEGISLATIVE DOCUMENTS IN DIGITAL LIBRARIES USING LINKED DATA

María Hallo¹, Sergio Luján-Mora², and Alejandro Mate³

¹ *Department of Computer Science, National Polytechnic School, Quito (ECUADOR)*

² *Visiting teacher at the National Polytechnic School, Quito (ECUADOR)*

^{2,3} *Department of Software and Computing Systems, University of Alicante, Alicante, Spain*
maria.hallo@epn.edu.ec, sergio.lujan@ua.es, amate@dlsi.ua.es

Abstract

Many countries have provided online access to some types of legislative documents by subject, keywords or date. Nevertheless, the possibility of querying historical versions of the documents is usually an uncommon feature. The dispersion of laws and other legislative documents and their continuous changes make difficult the generation and querying of valid legislative information at a given date. Furthermore, the ripple effect of modifications such as updates, insertions or derogations affecting the entire body of a law or part of it is not always visible for the citizens who are looking for legislative information.

Some issues related to change management of legislative documents can be identified: how to apply the history of changes to a version of a legislative document to obtain a new version, and what type of data model might be better to satisfy temporal queries, to store new versions of documents or to obtain them dynamically. The access to all versions of a document and its fragments is important in legislative queries to be sure which law was in force to apply when a case happened.

Law documents are produced and stored in information systems with different data models to access and retrieve information about them in a large-scale manner, but most of them do not have law change management functions. Web standards, such as XML, XSLT and RDF, facilitate the separation between content, presentation and metadata, thus contributing to a better annotation and exploitation of information from these documents and their fragments to improve the historical queries and the version generation of legislative documents.

This paper presents a proposal of a data model for storage and retrieval of different versions of legislative documents using Linked Data, a method of publishing structured interlinked data, for managing relations between legislative documents and its changes. Document structures, changes to legislation, metadata, requirements of historical queries are analyzed in this work. Furthermore, the proposed model facilitates historical querying of legislative documents and consolidation procedures, allowing update relationships between documents and fragments without changes on the original documents. The model has been tested with Ecuadorian laws, but it could be used for law systems of other countries because the model is independent of the legislative framework.

Keywords: Digital libraries, Linked Data, data models, legislative documents.

1 INTRODUCTION

Legislative documents can be classified into laws, decrees, resolutions and agreements. They are made in different official bodies and published in Official State Gazette such as: BOE in Spain, State Gazette in Bulgaria, the Gazette in United Kingdom, Official Gazette in Ecuador and so on.

Many countries have provided online access to some of the types of legal documents through information systems that allow searches by subject, keyword or chronological accesses. One of the most important requirements is querying enacted and consolidated versions of valid laws at certain

date. In order to facilitate the self-description, comparison, integration of documents and special queries some law systems use standards such as: XML (Extensible Markup Language), XSLT (Extensible Stylesheet Language Transformations), RDF (Resource Description Framework) [1,2].

However, recent experiences using Linked Data and Semantic Web with legislative information in England, Italy, Netherlands, Spain, and others [3,4] have shown that the updating process is slow, manual consolidations of versions are difficult, and there is a lot of outdated information.

In Ecuador and in other countries, there are legal systems maintained by private companies and official websites. The National Assembly of Ecuador begun to publish laws on its website after 2009 without relations between documents. In Ecuador, periodically consolidations of common laws are made and are published after approval by the National Assembly.

This paper presents a data model to meet the needs of historical queries and version generation of valid legislative documents on a given date allowing updates of relationships between documents without modifications on the original documents. The proposed model supports the consolidation of legislative documents developing consolidated documents integrated applying the corresponding changes. The model could be applied in several countries to their legal systems because it is independent of the legislative framework.

2 GENERAL CONCEPTS

The proposed model to store and retrieval legislative documents is based on the Linked Data principles for publishing and interlinking structured data on the Web. In addition, the model use metadata, information about the data, for making easier finding and working on legislative documents and their fragments.

2.1 Linked Data

Linked Data is a method of publishing data on the Web, using standards such as: RDF, XML, XSLT. RDF is a data model for describing resources on the Web. It models statements about resources in the form of triples subject (resource), predicate (property), object (value) building a graph. As such, RDF is suited to knowledge representation. RDF has been used to develop semantic models for legislation.

In 2006, Tim Berners-Lee presented the design principles of Linked Data [2]:

- Use URIs (Uniform Resource Identifiers) as resource names.
- Use HTTP URIs so that people can find these names.
- When someone searches for a URI should find useful information using RDF and SPARQL standards.
- Include links to other URIs so that people can find more related resources.

These principles are useful to share information in a way that can be read automatically by computers. This enables data from different sources being connected and queried.

2.2 Metadata

Metadata has been described as data about data. Metadata can be obtained from controlled vocabularies as ontologies, thesauri or uncontrolled vocabularies from particular designs. Metadata may be defined at the level of items or collections and may be embedded in the described object or defined outside with pointers to the objects they describe. There are multiple metadata standards, many of them for the same purpose with defined patterns of interaction allowing interoperability between them.

Definitions and criteria for comparing diagrams are useful for reusing vocabularies and are presented in the book "Introduction to metadata" edited by Mirtha Backa [5]. This proposal will reuse existing vocabularies with few additions to the amendments of laws, as specified in section 3.6.

3 DETAILS OF THE PROPOSAL

In this work, we suggest the use of a tree data model for legislative documents stored in XML format and a graph data model represented in RDF for metadata of legislative documents and links relating modifying and modified documents. The changes are produced generally from fragments of new legislation referencing fragments of old legislation. The tree data models fit well when component interconnectivity is a key feature for the legislative change management systems. The database system used is Virtuoso¹, a database management system supporting relational, XML and RDF data. The dynamic generation of versions is made applying the valid amendments until a given date to the original documents. The details of the model are shown in the following sections.

3.1 Content of Legislative Documents

The legislative documents may be of different types: laws, decrees, resolutions and so on. Each one has its own hierarchical structure and could have appropriated tags in XML.

A law in the case of Ecuador has the following parts:

- Heading
- Preamble
- Body
- Provision
- Annexes

The Body is composed of Titles, Chapters, Sections, Articles and Paragraphs, but not always we found all the components in a law. In some cases titles, chapters or sections are omitted.

Metadata are developed to describe the content of legislative documents, fragments and change links.

3.2 Metadata on Legislative Documents

Metadata of legislative documents and their changes will be represented in RDF in our proposal using the LegX ontology. This helps to dynamically query the metadata of a legislative item as it was enacted, its contents and history of their amendments on a given date. For specification of official versions the FRBR (Functional Requirements for Bibliographic Records) vocabulary [6] is used.

FRBR represents the products described in bibliographic records as work (law), expression (versions), manifestation (formats) and item (copy). Associating legislative information to FRBR vocabulary the work is a legislative document, it has expressions (versions) and each expression has manifestations (formats). Each manifestation can have items (copies).

Each work has a number of expressions that represent a different version of the legislation, each one with its respective date of enactment represented with the Dublin Core term: dct:issue (date of formal issuance of the resource). Dublin Core is a metadata model for use in resource description.

The FRBR vocabulary is used for structuring information about legislative documents. For the definition of links we used a vocabulary similar to Mets [7]. Mets is a metadata standard for encoding and transmission of digital objects.

The concepts and relations between the legislative documents, fragments and amendments are modelled in the proposed LegX (Legislative) ontology and it is represented in Figure 1. The terms legX: Changes and the properties changeFrom and changeTo pointing to the modified and modifying legislative documents are added to the proposal model. Moreover, metadata about global documents, fragments and amendments must be defined in each case using any of the existing metadata standards oriented to legislative systems such as CenMetalex o Akoma-Ntoso². CenMetalex [8] defines a machine readable representation (in XML format) of parliamentary and legislative documents. Moreover, the standard suggests a naming convention for URI-based identifiers for all structural elements of a legal document.

¹ Virtuoso Open-Source edition: <http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/>

² Akoma Ntoso XML for parliamentary, legislative & judiciary documents: <http://www.akomantoso.org/>

3.3 Identifiers

Similarly to the proposal of [legislation.gov.uk](http://www.legislation.gov.uk)³, URIs are used as identifier for legislative documents and fragments and also for each change link. URIs must be defined hierarchically showing the parts and version dates. Following is described a proposed URI structure for each component:

- The URI structure for legislative documents (work) is:
{domain/id/publication-year/publication-number/{legislative-item(promulgation-date)}}
Legislative-item {doc-type|title|chapter|section|article}
Doc-type {law|decree..}
- The URI structure for documents-versions (expression) is:
{domain/publication-year/law-id/{legislative-item/ver(version-date)}}
Legislative-item {doc-type|title|chapter|section|article}
Doc-type {law|decree..}
- The URI structure for documents-formats (manifestations) is:
{domain/publication-year/law-id/{legislative-item/ver(version-date)}/nom_file.extension}
Legislative-item {doc-type|title|chapter|section|article}
Doc-type {law|decree..}
- The URI structure for documents-formats-copies (items) is:
{domain/publication-year/law-id/{legislative-item/ver(version-date)}/nom_file.extension/numcopy}
Legislative-item {doc-type|title|chapter|section|article}
Doc-type {law|decree..}
- The URI structure for amendments links is:
{domain/legX/link/link-number/validity-date}

3.4 Temporal modeling of legislative documents and fragments

There have been several studies to define time dimensions related to legislative documents and applications. Boer et al. [9] define the following time data: publication date, enactment date, start of validity date that in some cases matches with the publication date, repeal date (the end of the validity), effectiveness date, applicability date. Moreover, Boer et al. define the state of a document as active or inactive. In addition, Grandi et al. [10] propose timestamps for multiversion XML documents: start and end of validity, and effectiveness of transactions.

The time dimensions may be associated with the works (legislative documents) and expressions (versions of legislative documents) being inherited by the manifestations (formats). It is proposed to use the following data: publishing date, start of validity date, end of validity date (repeal), start of effectiveness date, end of effectiveness date for versions of documents and fragments.

³ UK Legislation website by National Archives: [http: www.legislation.gov.uk](http://www.legislation.gov.uk)

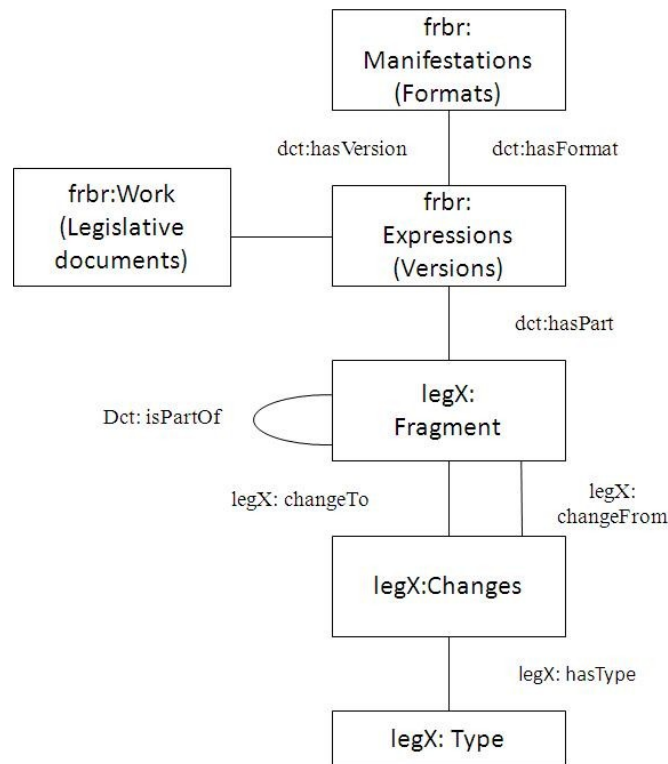


Figure 1. LegX ontology

3.5 Changes in legislation

Acts and regulation have continuous changes. Changes must be recorded and applied to old laws. The changes are producing from fragments of new legislation referencing fragments of old legislation. Different types of changes could be produced: inserted, updated, repealed. A common request is to view summaries of significant changes to specific acts or regulations. For applying changes we define links with the following attributes:

- ChangeDate: is the date on which the amendment applies to the target document.
- From: URI of a modifying fragment.
- To: URI of a modified fragment.
- Type of change: Inserted, deleted, repealed, updated.
- Description of change: Description of the reasons for change.

The changes applied to the original document generate a new version dynamically.

3.6 Description RDF / XML

For a description of legislative documents, versions, modifications and fragments we propose reuse the vocabularies shown in Table 1. These vocabularies have the concepts needed to our model.

Table 1. Vocabularies used in the data model

Vocabulary	Prefix	Namespace
RDF Schema	Rdfs	http://www.w3.org/2000/01/rdf-schema#
Dublin Core	dct	http://purl.org/dc/terms/
FRBR	frbr	http://purl.org/vocab/frbr/core#
Legislative	legX	http://mydomain/gob/ec/legX

The legX ontology was developed to represent the FRBR elements and links relating the changes. Versioned documents (expressions) point to its manifestations (formats) with the Dublin Core element `dct:hasFormat`. A legislative item of type work points to its parent item with `dct:isPartOf`. A legislative item of type work points to its component items with `dct:hasPart`. Each legislative document is linked with its versions with `dct:hasVersion`: with a reverse link `dct:isVersionOf`, it's shown in Figure.1.

Following we present a fragment of RDF/XML format for metadata of a legislative document from Ecuador defined at work (document) level, showing the relation to two consolidated versions. Each element has his own tag in XML.

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:dct="http://purl.org/dc/terms/"
xmlns:frbr="http://purl.org/vocab/frbr/core#"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
<frbr:Work rdf:about="http://www.midominio.gob.ec/id/1960/10">
  <dct:title> Ley de Propiedad Horizontal </dct:title>
  <dct:description> Regula la división y organización de diversos inmuebles, como resultado de la segregación de un edificio o de un terreno común.
  </dct:description>
  <dct:issue> 16/03/1960 </dct:issue>
  <dct:hasVersion> rdf:resource="http://www.midominio.gob.ec/1960/10/ver/16-03-1960"
  </dct:hasVersion>
  <dct:hasVersion>rdf:resource="http://www.midominio.gob.ec/1960/10/ver/06-10-2005"
  </dct:hasVersion>
  .....
</rdf:RDF>
```

Metadata about changes are represented by links in RDF/XML vocabulary. In the next example we show the metadata of a link relating the modifying and the modified norms identified in (link:from) and (link:to) tags.

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dct="http://purl.org/dc/terms/"
  xmlns:link="http://www.midominio.gob.ec/legx/link">
<rdf:description rdf:about="http://www.midominio.gob.ec/legX/link/link1/27-09-2011">
<link:from rdf:resource="http://www.midominio.gob.ec/2011/27/art/2/p/1/ver/27-09-2011"/>
  <dct:title> Link1 </dct:title>
  <link:to rdf:resource="http://www.midominio.gob.ec/1960/10/art/2/ver/06-10-2005"/>
  <link:type rdf:resource="http://purl.org/leg/update"/>
  <link:valid-start>"27-09-2011"</link:valid-start>
</link>
</rdf:description>
</rdf:RDF>
```

3.7 Consolidation

Consolidation is the process of applying to a legislative document all changes made over time. In order to have official resulting documents, they must obtain the approval of the competent bodies as in the case of the National Assembly in Ecuador.

As an aid to consolidation process for each legislative document or fragment it is possible to query the changes made to it at a given date or in a range of dates.

The proposed model allows handling versions of official documents and unofficial consolidated versions dynamically generated supporting the official consolidations or being a reference for system users.

3.8 Queries

Most of the legislative information systems allow to query the legislative document valid at a specific date and to get a full-text search upon the legislative document. Moreover, the historical evolution of a document is also covered by several systems, but, the historical evolution of a fragment is less supported. In the same way, information about amending and amended acts of a rule is not always considered and the queries related with validity and efficacy is less supported [11]. In general, the Linked Data oriented proposals support a wider set of queries using SPARQL [12].

Our proposed model allows the following queries:

- a) A valid legal document to a certain date.
- b) Historical development of a legislative document with access to all the approved versions.
- c) A piece of legislation and its amendments.
- d) Laws modified by a law.
- e) Validity range of a legislative document.
- f) Validity range of a piece of legislation.
- g) Efficacy range of a legislative document.
- h) Efficacy range of a piece of legislation.

3.9 Advantages of the proposed model

With this proposal, metadata are separated from the legislative documents and can be expanded (maintained independently) adding new metadata without changes in the original document. This model allows generating a new version of a legislative document dynamically applying the changes to the original document. Moreover, the model allows linking in the future the documents and the modifications to other resources as the proponents, the voting results, etc. [13]. This model has been implemented using Virtuoso, software with capabilities to manage relational data, XML fields and RDF structures.

4. CONCLUSIONS AND FUTURE WORKS

The proposed model integrates the information of metadata from legislative documents and amendments using RDF. Moreover, it's possible to add another relationships in RDF in order to cover the process of issuing laws and future changes. In addition, metadata are separated from the documents and can be expanded independently. Furthermore, the proposed model supports the consolidation of legislative documents developing integrated documents with the corresponding amendments and could be applied in several countries to their legal systems. For the initial loading of metadata in RDF, it is necessary to define an extraction process from XML fields, according to the type of data sources and their corresponding metadata. In the future, another kind of legislative process should be analyzed to integrate new documents and relationships in the model. In addition, tests of performance will be addressed.

5 ACKNOWLEDGMENTS

This work has been partially supported by the Prometeo Project by SENESCYT, Ecuadorian Government.

REFERENCES

- [1] Peri, A. (2009).SIDRA:XML en la gestión y explotación de la documentación jurídica. SCIRE. 15, pp.111-124.
- [2] Martínez, M., et al. (2003). Relationship-based dynamic versioning of evolving legal documents. Web-knowledge Management and Decision Support, Lecture Notes on Artificial Intelligence. 2543, pp. 298-314.
- [3] Sheridan, J. and Tension, J. (2010). Linking UK Government Data, Statistics, ACM Press, North Carolina, pp.1-4.
- [4] Hoekstra, R.(2011).The MetaLex Document Server, in VoxPopuLLI. Available from: <http://blog.law.cornell.edu/voxpath/2011/10/25/the-metalex-document-server/> .[Accessed: 2 Jan 2015].
- [5] Baca M., Gill T., Gilliland A.J., Whalen M. and Woodley M.(2008). Introduction to Metadata, Getty Research Institute, Getty Publications, Los Angeles.
- [6] IFLA Study Group. (1997). Functional Requirements for Bibliographic Records, IFLA Series on Bibliographic Control, 19, Munich: K.G. Saur Verlag.
- [7] Cundiff, M. V. (2004). An introduction to the metadata encoding and transmission standard (METS). Library Hi Tech, 22(1), pp. 52-62.
- [8] Palmirani,M., Cervone, L. and Vitali, F.(2009). Legal metadata interchange framework to match CEN metalex. In *Proceedings of the 12th International Conference on Artificial Intelligence and Law* (ICAIL '09). ACM, New York, NY, USA, pp. 232-233.
- [9] Boer, A., Hoekstra, R., Winkels, R., van Engers, T., Breebaart, M. (2004). Time and Versions in METALex XML. In: *Proceeding of the Workshop on Legislative XML*. Kobaek Strand.
- [10] Grandi, et ál. (2005) Temporal Modelling and Management of Normative Documents in XML Format. *Data & Knowledge Engineering*, 54, pp. 327–354.

- [11] Hallo, M. Martínez, M., De la Fuente, P. (2013). Data models for version management of legislative documents, *Journal of Information Science*. 39(4), pp. 557-572.
- [12] Sartor, G., Palmirani, M., Francesconi, E., Biasiotti, M. (ed). (2011). *Legislative XML for the Semantic Web*, 1st Ed. Netherlands: Springer.
- [13] Hallo, M., Luján-Mora, S. Trujillo, J. (2014). Transforming Library Catalogs into Linked Data, *ICER2014 Proceedings*, pp. 1845-1853.