A12
Monica Palmirani

Legislative XML: Principles and Technical Tools
Revised by Fabio Vitali\textsuperscript{1}, Department of Computer Science, University of Bologna
Translated in Spanish version by German Gomez Ventura, European University Institute\textsuperscript{2}

This document was commissioned by the Inter-American Development Bank (IDB) to
the University of Bologna - CIRSFID and the Department of Computer Science - within
the framework of the Regional Technical Cooperation "Connected Parliaments: The
Impact of New Technologies in the Transformation of the Legislative Branch" (RG-
T1776), supported by the Italian Trust Fund for Information and Communication
Technology for Development in an effort to systematize information and knowledge
about the potential of new technologies in the modernization of the legislative branch.

\textsuperscript{1} Fabio Vitali is associate professor of Computer Science at University of Bologna, Computer Science
Department.
\textsuperscript{2} German Gomez Ventura translated in Spanish the version of the document, nevertheless for
permitting to the native Spanish audience to better follow the examples, the context of the
exemplifications was completely customized.
# TABLE OF CONTENTS

**INTRODUCTION** ............................................................................................................... 7

**PART I THE OPPORTUNITIES AND RISKS OF LEGISLATIVE XML** ................. 9

1. **INTRODUCTION TO XML** ..................................................................................... 9
   1.1 WHAT IS XML? .................................................................................................. 9
   1.2 HOW TO USE XML ......................................................................................... 15
   1.3 USEFULNESS OF XML .................................................................................. 16
   1.4 LEGISLATIVE XML .......................................................................................... 19
   1.5 COSTS ASSOCIATED WITH LEGISLATIVE XML ........................................ 23
      1.5.1 Tagging Time ............................................................................................. 25
      1.5.2 Software Costs .......................................................................................... 26

2. **ADVANTAGES OF USING LEGISLATIVE XML** ............................................. 27
   2.1 BENEFITS FOR PARLIAMENTARY WORK ..................................................... 27
      2.1.1 Accessibility ............................................................................................... 27
      2.1.2 Transparency ............................................................................................. 27
      2.1.3 Simplification ............................................................................................... 27
      2.1.4 Independence from Proprietary Platforms ................................................. 28
      2.1.5 Multi-channel Publication .......................................................................... 28
      2.1.6 Reuse ........................................................................................................... 28
      2.1.7 Interoperability ........................................................................................... 28
      2.1.8 Comparative Analysis of Legislation ......................................................... 29
   2.2 BENEFITS OF APPLYING XML ........................................................................ 29
      2.2.1 Texts are constantly updated ..................................................................... 29
      2.2.2 Simulating Regulation and Quality Control ............................................... 29
      2.2.3 Semantic Searching ..................................................................................... 29
      2.2.4 Long-Term Conservation .......................................................................... 30
      2.2.5 Self-Contained Knowledge ......................................................................... 30

3. **RISKS INVOLVED IN USING XML IN THE LEGISLATIVE PROCESS** ...... 31
   3.1 POLITICAL RISKS ............................................................................................ 31
      3.1.1 Political commitment .................................................................................. 31
      3.1.2 Interoperability with other agencies ......................................................... 31
      3.1.3 Long-term economic hedge ....................................................................... 31
      3.1.4 Legislative drafting guidelines ................................................................... 32
   3.2 ORGANIZATION AND MANAGEMENT RISKS .............................................. 32
      3.2.1 End User Involvement ................................................................................ 32
      3.2.2 Office Reorganization ................................................................................ 32
      3.2.3 Personnel Training and Assistance ............................................................. 32
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.2.4 Project Management</td>
<td>33</td>
</tr>
<tr>
<td>3.3 TECHNOLOGICAL RISKS</td>
<td>33</td>
</tr>
<tr>
<td>3.3.1 Proprietary XML Standards</td>
<td>33</td>
</tr>
<tr>
<td>3.3.2 Maintaining the Standard over Time</td>
<td>33</td>
</tr>
<tr>
<td>3.3.3 ConvertingExisting Databases</td>
<td>34</td>
</tr>
<tr>
<td>3.3.4 Markup Time</td>
<td>34</td>
</tr>
<tr>
<td>3.3.5 The Choice of Tools</td>
<td>34</td>
</tr>
<tr>
<td>PART II HOW TO USE LEGISLATIVE XML</td>
<td>37</td>
</tr>
<tr>
<td>4. THE STATE OF THE ART IN LEGISLATIVE XML</td>
<td>37</td>
</tr>
<tr>
<td>5. MANAGING LEGISLATIVE DOCUMENTS</td>
<td>41</td>
</tr>
<tr>
<td>5.1 MODELING DOCUMENTS</td>
<td>41</td>
</tr>
<tr>
<td>5.2 DESCRIPTIVE AND PRESCRIPTIVE</td>
<td>43</td>
</tr>
<tr>
<td>5.3 GLIN AND LEGISLATIVE XML</td>
<td>44</td>
</tr>
<tr>
<td>6. LEGAL DOCUMENT VALUE OVER TIME</td>
<td>45</td>
</tr>
<tr>
<td>6.1 AUTHENTICITY, INTEGRITY, VALIDITY, AND PERSISTENCY</td>
<td>46</td>
</tr>
<tr>
<td>6.2 LEGAL VALIDITY WITH DIGITAL SIGNATURES</td>
<td>50</td>
</tr>
<tr>
<td>7. AKOMA NTOSO</td>
<td>53</td>
</tr>
<tr>
<td>7.1 DEFINITION OF A LEGAL DOCUMENT</td>
<td>54</td>
</tr>
<tr>
<td>7.2 TYPES OF LEGISLATIVE DOCUMENTS</td>
<td>55</td>
</tr>
<tr>
<td>7.2.1 Legislative Documents: Bills and Acts</td>
<td>55</td>
</tr>
<tr>
<td>7.2.2 Amendments</td>
<td>56</td>
</tr>
<tr>
<td>7.2.3 Documents</td>
<td>57</td>
</tr>
<tr>
<td>7.2.4 Debate Documents</td>
<td>58</td>
</tr>
<tr>
<td>7.2.5 Judgments</td>
<td>62</td>
</tr>
<tr>
<td>7.2.6 Document Collections</td>
<td>63</td>
</tr>
<tr>
<td>7.3 LEGAL DOCUMENTS: AN OVERVIEW</td>
<td>65</td>
</tr>
<tr>
<td>7.3.1 The Tripartite Structure of a Legal Document</td>
<td>65</td>
</tr>
<tr>
<td>7.4 NAMING LEGISLATIVE DOCUMENTS</td>
<td>68</td>
</tr>
<tr>
<td>7.5 NAMING LEGISLATIVE RESOURCES: DEFINING THE URI</td>
<td>69</td>
</tr>
<tr>
<td>7.6 LEGISLATIVE DOCUMENTS METADATA</td>
<td>70</td>
</tr>
<tr>
<td>7.6.1 Metadata in Akoma Ntoso</td>
<td>74</td>
</tr>
<tr>
<td>PART III TOOLS</td>
<td>79</td>
</tr>
<tr>
<td>8. SPECIALIZED LEGISLATIVE XML EDITORS</td>
<td>81</td>
</tr>
<tr>
<td>8.1 TEXT PROCESSOR EDITORS</td>
<td>81</td>
</tr>
<tr>
<td>8.1.1 Norma-Editor</td>
<td>84</td>
</tr>
<tr>
<td>8.1.2 Bungeni-Editor</td>
<td>90</td>
</tr>
<tr>
<td>8.2 NATIVE XML EDITORS</td>
<td>93</td>
</tr>
</tbody>
</table>
8.2.1  *xmLeges* ................................................................. 93
8.2.2  *LegisPro* ................................................................. 94
8.3   *Web Editor* ............................................................... 95

9. DATABASE MANAGEMENT SYSTEMS FOR LEGISLATIVE XML DOCUMENTS ................................................................. 99
   9.1   *Norma-System* .......................................................... 99
   9.2   *Bungeni* ................................................................. 101

10. LINK SOLVER ................................................................. 105

PART IV MARKUP METHODOLOGY IN LEGISLATIVE XML ................................................................. 109

11. LEGAL ANALYSIS OF THE DOCUMENT ................................................................. 109
   11.1  Determining the Type of Document ........................................ 109
   11.2  Determining the Parts of the Document .................................... 109
   11.3  Determining the Identifying Elements of the Document .......... 112
   11.4  Determining the Document Version ....................................... 112
   11.5  Isolating each Main Legal Block in the Document ................. 113

12. HOW TO MARKUP THE LEGISLATIVE DOCUMENT STRUCTURE IN XML ................................................................. 117
   12.1  Determining the Preface of the Document ............................ 118
   12.2  Determining the Preamble of the Document .......................... 120
   12.3  Determining the Normative Structure of the Document .......... 121
   12.4  Determining the End Part of the Document .......................... 123

13. HOW TO MARKUP THE LEGAL METADATA FOR LEGISLATIVE DOCUMENTS IN XML ................................................................. 125
   13.1  Determining the Identification Metadata of the Document ....... 125
   13.2  Detecting References .................................................... 126
   13.3  Determining the Referencing Metadata .................................. 128
   13.4  Representing Metadata for Temporal Events ........................ 128
   13.5  Identifying Notes by the Author or Official Authority ............ 128

14. HOW TO MARKUP THE LEGAL SEMANTICS OF LEGISLATIVE DOCUMENTS IN XML ................................................................. 131
   14.1  Detecting Semantic Elements of the Text ............................ 131
   14.2  Analysing the Provisions ............................................... 133
   14.3  How to Markup Consolidated Legislative Documents in XML ... 135
   14.4  How to Use the XML Files ............................................ 138

15. CONCLUSIONS ................................................................. 141
Table of Contents

GLOSSARY ..................................................................................................................... 142
ANNEXES ..................................................................................................................... 147
WEBSITE REFERENCES ............................................................................................. 159
REFERENCES ............................................................................................................. 160
INTRODUCTION

This document, commissioned by the Inter-American Development Bank (IDB) and developed by the University of Bologna\textsuperscript{4}, presents advantages, opportunities, and risks involved in applying Legislative XML to a parliamentary workflow. It is aimed at a diverse audience of experts involved in a parliamentary law-making system, and in particular the following five groups of end users:

**Parliamentary members, legislators, and other decision-makers.** The goal is to enable this group to have an appreciation of the Legislative XML format and tools - how they work and what they can achieve in the work of the legislature, in policymaking as well as in communication with voters and citizens (Part I).

**Legislative officers and senior legislative officials.** It is important for people in this group to understand the basic concepts of Legislative XML standards and the way that bespoke tools can help legislative officers streamline their work. The goal of using XML is to optimize the legislative process, improve interoperability, minimize costs, and maximize effectiveness; thus improving the quality and accuracy of the legislative process (Parts I and II).

**Drafters and legislative clerks.** These professionals must be able to combine legal drafting techniques (legislative/technical manuals) with best practice in Legislative XML markup, analyzing legislative documents from different perspectives to capture the semantic parts of the text, while maintaining the linguistic aspects of drafting legislation. It is essential for this group to understand which XML conversion tools are available on the market and the benefits such tools can bring to their work (Parts III and IV).

---

\textsuperscript{4} CIRSFID and the Department of Computer Science, University of Bologna.
ICT managers and staff responsible for the technology used to create and maintain legislative documents. The requirement here is for technical understanding of Legislative XML, its strengths and weaknesses, with the necessary skill-set to build and support Legislative XML applications, while understanding the technical and organizational challenges involved in the application of XML to a legislative environment (Parts II and III).

ICT developers. This is the group entrusted with creating, implementing, and supporting technical applications used to produce and maintain legislative documents. People in this group need to understand key technical specifications for the development of ICT solutions that are compliant with the Legislative XML approach. They can also draw inspiration from the tools presented here (Parts II, III, and IV).

The way the document is organized reflects the intended audience:

- Part I offers a general introduction to the benefits and risks of Legislative XML;
- Part II introduces Legislative XML features with particular attention to the Akoma Ntoso standard;
- Part III presents XML markup support tools; and
- Part IV shows how to apply Akoma Ntoso and mark up a legal document in a way that respects the meaning of each part.
PART I
THE OPPORTUNITIES AND RISKS OF LEGISLATIVE XML

1. INTRODUCTION TO XML

1.1 What Is XML?

XML (Extensible Markup Language) is a markup language developed by W3C\(^5\) (www.w3.org) as an open standard, and which has been extensively adopted worldwide to represent data and documents.

The main idea behind XML is to enclose text within annotation elements, or tags\(^6\), which serve to label the text. Just as on paper we use a colored marker to highlight the parts of a text that we want to single out, in XML we use bookmarks to attribute a certain meaning to selected parts of a text. In the following example, the docType tag identifies the type of legislative document, the docDate tag provides the date of enactment, and the docTitle tag provides the full title of the legislation. Thus, for example, we can ask the system to find all 2010 laws that contain the words “collective property” in the title.

```xml
<preface>
  <subdivision>
    <docProponent>Republic of Panama</docProponent>
    <docProponent>Ministry of Agricultural Development
  </subdivision>
</preface>
```

\(^5\) http://www.w3.org/
\(^6\) The tag is composed of two parts: an opening tag (<docTitle>) and a closing tag (</docTitle>). The name of the opening and closing tags are the same. Note that they are case sensitive (doctitle is different to docTitle). The symbolic characters <, >, / are used to isolate the tag from other parts of the markup. The content is enclosed within the opening and closing tags. For example: <docTitle> this is a text </docTitle>.
XML also permits the addition of metadata about the document, which combined with the structure of the text, adds information not directly included in the narrative. At the same time, XML allows the performance of a semantic query such as: “show me all the documents that talk about <Agrarian Reform> of the articles in force in 2009”. This is possible even if those words are not included in the text of the article:

```
<classification source="#library"/>
<keyword id="keyword1" value="agrarianReform" showAs="Agrarian Reform" dictionary="Tesauro"/>
```

```
<article id="art2" refersTo="#keyword1">
<num>Article 2:</num>
<clause id="art2-cls1">
<p>
The collective title application must be accompanied by the following documents and information:
1. The holder’s plan or sketch of the area.
2. A certificate issued by the General of the Contaloría Republic with a census of the population of the community.
3. Certification by the Indigenous Policy Directorate of the Ministry of Government and Justice proving the existence of the applicant community or communities.
4. Name and general information about the authority or person making the request on behalf of those bordering
```

Figure 1 – XML fragment.
indigenous communities.</p>
</content>
</clause>
</article>

Figure 2 – Classification of the text.

- **XML is an open-source format for data and documents**, which means that no one company owns the copyright. The format is documented and shared within the IT community, and is delivered through an open-source license.

- **XML is machine readable**. The example presented in the tables below shows how XML documents can be understood by both humans and computers.

- **XML is technology-neutral**. XML can be used with any open-source program or simply with a non-proprietary editor. This way, the parliament will not be tied to any one product to process its data and documents, and investment can be made in a technology that so far does not depend on the software applications on the market. This implies that we will no longer find that we cannot open a document just because the application we use is outdated and its use depends on updating or buying a new product on the market.

- **XML identifies the purpose and content of the information** and can be used for multiple purposes such as searching and preservation as well as presentation on multiple devices. It therefore has more capabilities than HTML.

- **XML is machine-readable**. Unlike PDF, the semantics of each part of the XML file can be understood by a computer, which is able to manipulate any fragment of the document. In fact, the tags help the computer to understand the meaning of the marker-defined values. For example, if you enclose an article’s contents in the “article” tag, the computer will recognize all parts of the text as having a normative value.

- **XML is hierarchical**, rigorous, extensible, highly accurate, and flexible. Because XML is a language of languages, each community can build its own vocabulary of tags depending on the problem to be
solved. On the legislative front, there are families of Legal XML standards that have emerged to capture the peculiarities of legal and legislative documents. At the same time, XML is rigorous, because like any natural language, it uses lexicon and syntax (i.e., a grammar) to define its own rules. This way, it defines a behavioral rule for a tag (requiring, for example, that all paragraphs be numbered). It is impossible for the user to infringe this rule because the program will check if the XML is correct, and detect any errors. This requires that the user strictly observes the standardization and harmonization rules in such a way that the resulting document is more accurate.

- **XML ensures the long term preservation and accessibility of the document.** Because XML is an open-source, *open* markup language, it is an optimum data format for storing information for the long term, as is so often necessary when dealing with legislative and legal information. The format is designed so that we can access legal documents a hundred years from now without having to worry about finding software that can read them.
Figure 3 – XML document example: end-user visualization and code.
XML was born in 1996 in an attempt to overcome the limitations of HTML and in an effort to make the language of markers more rigorous. As compared to other formats, XML has a high expressive value, and once a document is converted into XML, it can easily be converted into any other format, as shown in the pyramid figure below.

The higher up in the pyramid, the more valuable the information and its potential for reuse, even if it takes more time to convert that information from paper to XML.
1.2 How to Use XML

XML is used in a document by placing markers (tags) around various components of the text, such as the title, date of introduction, part, article, conclusion, etc. Firstly, it is necessary to define a vocabulary of tags and rules specifying the meaning of the tags, their format, and how they relate to one other. These rules are defined by special instruments that define templates, of which there are two types: Document Type Definitions (DTDs) and XML schemas. Both are techniques used to define a glossary of tags, and to set out the grammatical rules with which to use them. A schema acts as a sort of guide concerning how to use the vocabulary that a given community shares when checking text for quality, harmonization and interoperability.

Having studied the schema (DTD or XML), the user can apply the tag to the text manually or by using specialized editors.

A specialized editor will check the XML file for correctness to make sure that the tags are well formed and that the XML method was used properly. It checks only for well-formedness, not content.
1.3 Usefulness of XML

What does XML do that other data formats or structures database don’t?
**Introduction to XML**

**Machine readability.** Firstly, XML is machine readable: the semantics conveyed by text markers is intelligible to the computer, which can then process the legal information for different purposes in a very rapid manner. Databases would force and delimit the document according to patterns that the programmer had decided in advance, and these patterns would not be able to convey the richness of the information contained in the document itself. XML makes it possible to describe information accurately (down to a single word), giving an expert in the subject complete freedom to decide how to achieve the purpose in hand. Finally, XML cannot alter the document’s content, so you know that the document will be legally valid over time.

Several more uses of XML are possible in a parliamentary context:

**Links between norms.** It is possible to link all the bills laid down in parliamentary debates and accurately determine which parts of the bill originated from a particular member of parliament, thus establishing a link between information in session reports and final bills. It is an excellent tool to promote transparency regarding what goes on in parliament.

**Better search capability.** Texts can be searched in a very granular way, no longer forced to follow a fixed model such as the former database method, but able to use data in the tags and the text together. In addition, several search channels can be combined: structured data, text searches, or searches on tags.

**Multiple platforms.** The same XML document can be used to publish the content it contains across multiple platforms: the Internet, mobile, online TV, kiosks, online books, etc. Content can thus be delivered through multiple media without having to modify the XML document.

**Dialogue between systems.** Dialogue between different systems and organizations are enabled via the XML language, which is an excellent exchange format. This promotes interoperability among institutions and facilitates exchange of data.

**Diverse applications.** A document can be opened using different software applications, rather than just one, so there is more freedom in the organization of parliament, which no longer needs to be tied to any one
product. Above all, an XML document can be used by all network technologies that take XML as a key pillar: blog, wiki, social networking, RSS, etc.

**Reuse of information.** Information is reused for different purposes and at different stages in the legislative process: the agenda of a parliamentary session consists of a list of documents (proposals, written requests, bills to be debated, etc.) that have been previously deposited and marked up in XML. Much information, such as normative references, the proponents, and other important metadata, have already been marked up and classified. With XML, one can create a session agenda by extracting key information from existing documents; the agenda can be used in a stenographic transcript of the session’s proceedings; and, finally, the abstract or the draft can be a by-product of the stenographic transcript.

![Diagram showing the legislative cycle with information being reused](image)

**Figure 7 – Example of how information can be reused in the legislative cycle.**