Managing semantics in XML vocabularies: an experience in the legal and legislative domain

Gioele Barabucci[1], Luca Cervone[2], Angelo Di Iorio[1],
Monica Palmirani[2], Silvio Peroni[1], Fabio Vitali[1]

[1] Department of Computer Science, University of Bologna, Bologna, Italy,
[2] CIRSFID, University of Bologna, Bologna, Italy
barabucc@cs.unibo.it, luca.cervone@unibo.it, diiorio@cs.unibo.it,
monica.palmirani@unibo.it, speroni@cs.unibo.it, fabio@cs.unibo.it

Abstract. Akoma Ntoso is an XML vocabulary for legal and legislative documents sponsored by the United Nations, initially for African Countries and subsequently for use in other world countries. The XML documents that represent legal and legislative resources in Akoma Ntoso contain a large quantity of elements and sections with concrete semantic information about the correct description and identification of the resource itself and the legal knowledge it contains. Such information is organized in many distinct conceptual layers, allowing for the contribution of different semantic information according to competencies and role in the workflow of the contributor.

This paper shows how the Akoma Ntoso standard expresses the independent conceptual layers of semantic information, and provides ontological structures on top of them. We also discuss how current Semantic Web technologies could be used on these layers to reason on the underlying legal texts.

As one of the main funding principles of Akoma Ntoso is the long-term preservation of legal documents and of their intended meaning, this paper also shows and justifies some design decisions that have been made in order allow future toolmakers to access the enclosed legal information without having to rely on current technology that may be long forgotten in the future decades.

Keywords: GRDDL, Legal and legislative resource, Legal reasoning, Parliamentary documents, XML vocabulary

1 Introduction

Akoma Ntoso is an XML vocabulary for legal and legislative documents whose primary objective is to provide semantic information on top of a received legal text. There are three key aspects of legal documents on which Akoma Ntoso focuses: identification of structures, references to other legal documents and storage of non-authoritative annotations. Structures are identified and marked up according to an XML vocabulary based on common patterns found in legal documents. References to legal documents across countries are made using a common naming convention based on URIs. Third-party annotations and interpretations (broadly called *metadata*) are stored using and ontologically sound approach compatible with Topic Maps [15], OWL [19] and GRDDL [4].

The XML documents created according to the Akoma Ntoso specifications use a layered structure where each layer addresses a single problem: the *text* layer provides a faithful representation of the original content of the legal text, the *structure* layer provides a hierarchical organization of the parts present in the text layers, the *metadata* layer associate information from the underlying layers with ontological information. Whenever this semantic information is the result of a subjective interpretation, Akoma Ntoso allows multiple and independent opinions to be stored in a formal way within the document, and used alternatively, cumulatively or compared to each other.

The layered structure of Akoma Ntoso is an attempt at balancing extensibility, needed to accommodate the specific needs of individual countries, with clarity and self-explanatoriness, both needed for the preservation of legal digital resources over time (even long spans of time, measured in decades or centuries). Both these aspect have been evaluated taking into account the fact that long preservation of Akoma Ntoso documents must be possible even without access to the extensive original documentation.

The same layered structure creates a strict separation between the content that has been approved by the body empowered by law to endorse it (data) and what has been added by other parties (metadata). This separation significantly helps the development of tools able to preserve and guarantee the authenticity of the processed legal document, favouring trust towards e-government initiatives. In fact, Akoma Ntoso XML documents can be managed in any step of the legislative or judiciary life cycle (for instance, in the publishing phase) without any modification to the received text.

While Akoma Ntoso imposes an (extensible) XML vocabulary, it does not prescribe the use of a particular ontology. Actually Akoma Ntoso defines a minimal and loose ontology based on few *Top Level Classes* (TLCs) e.g., Person, Role, Concept, etc. These classes are only generic groupings of instances: no particular property is defined for any of them. Inside an Akoma Ntoso document, a section of the metadata links pieces of text with the appropriate TLC instances, another section of the metadata combines these instances to create complex relations.

To perform elaborated computations on a document or on a collection of documents, more precise ontologies have to be used and linked with the provided metadata. For example, we may be interested in using the FRBR (Functional Requirements for Bibliographic Records) ontology to associate some of the document metadata describing legislative documents to FRBR concepts like Work or Expression of a Work. Another example is the representation of individual persons: instances of the TLC Person class may be associated to instances of the Person class of the FOAF (Friend of a Friend) ontology or to instances of the Creator class of the Dublin Core ontology. Akoma Ntoso allows the use of these and of future ontologies.

Even if one relies only on the bare knowledge provided by the Akoma Ntoso minimal ontology and by its document markup, there are many interesting queries that can be carried out using only the original document. Some of these queries can be expressed through XPath expressions [2], even though it is syntactically and semantically better to query documents metadata (extracted and stored in RDF statements) through SPARQL [1]. Another way to use the same semantic information (accompanied or not accompanied by an external ontology) it to use it as a knowledge base on top of which systems based on LKIF (Legal Knowledge Interchange Format) [5] or RuleML [6] can operate.

This paper is organized as follows. Section 2 illustrates the layered architecture of the Akoma Ntoso vocabulary. Section 3 describes the lower layers of the Akoma Ntoso document architecture where the legal text is described and the hooks used by upper layers are added. Section 4 shows how authoritative and non-authoritative legal knowledge (metadata) can be codified in an Akoma Ntoso document. Finally, Section 5 illustrates how current semantic tools can use the generic Akoma Ntoso ontology system to reason over legal documents.

Note: throughout this paper we will use the terms *legal text* and *legal resource* for legal texts as endorsed by an official authority; we will use the term *legal document* for their computer representation; we will use the terms *legal content* or *normative content* for pointing out the normative meaning of the text.

2 The Akoma Ntoso document architecture

The role of Akoma Ntoso is to mark up legal and legislative texts so that the legal knowledge and the legal structure of the text can be understood by current and future computer applications. This means that legal texts form the basis on which Akoma Ntoso documents exist. Akoma Ntoso documents add information on top of the actual text. The added information can itself be seen as composed of different stratified layers (text, structure, metadata, etc.) [12]. Akoma Ntoso clearly separates the legal text from these different levels of information but still allows higher layers to reference the underlying layers, thus building knowledge on top of other knowledge, with the content of the legal text acting as the base knowledge.

As with any technology that deals with legal resources, Akoma Ntoso has been designed to work on the original text without changing to it. Words and punctuation marks, but also other typographical symbols, are left untouched by the mark up process that transforms a plain-text legal document into an Akoma Ntoso-compliant XML document.

Additionally, Akoma Ntoso maintains a strict separation between data and metadata and provides an unambiguous definition of them as well as an operational distinction in authoriality: as such, data is any information that has been created or at least approved by the relevant legal author (for example the whole of the text of an act), while metadata is any information that was not present in the original version of the document as it was approved by the relevant legal author but was added editorially in a later moment of the production process (e.g., the issue number of the official gazette or, even, the page numbers in the printed version of the same act).

The distinction between data and metadata is not only a theoretical distinction, since the actual layers of markup in Akoma Ntoso, *text*, *structure* and *metadata*, are based on it.

Textual markup identifies, within the content of the legal documents, fragments that have a precise legal or referential meaning, e.g., concepts such as “this piece of text is a date”, “this piece of text is a legal reference” or “this piece of text contains the name of a party of the trial”. Structural markup identifies and organize the parts of the content that divide it into containers, and especially hierarchical containers: “this piece of text is an article” or “this piece of text is the title of an act”, “this piece of text is the background section of a judgement”, etc. Metadata markup adds knowledge generated

by an interpretation of the legal text performed by an human or mechanical agent: “the phrase the pre-existing Acts refers to Act 32 of 1989 and Act 2 of 1990”, or “the person cited in the minutes as Mr. Gidisu is really Mr. Joe Kwashie Gisidu, the only member of the current Ghanaian parliament with that name and elected to a seat in it since 2000”.

The analysis of the textual and structural information is quite straightforward and its results are rarely disputed. On the contrary, the analysis of more advanced concepts found in the legal text requires some experience and it is easy for different sources to disagree on the generated interpretation. For this reason Akoma Ntoso documents have exactly one textual and one structural layer in each documents while interpretation of the advanced concepts is stored as metadata, and Akoma Ntoso allows multiple metadata layers in the same document, each providing an interpretation by a different source.

Finally, each interpretation added by a specific actor can be linked to ontologies of legal concepts (e.g. date of enter into force as modelled by the LKIF-core ontology or the concept of the High Court of South Africa as modelled in an ontology about the judiciary system in Africa) by associating ABox assertions, described in the XML document and extractable in a more proper data model using GRDDL [4], to the general TBox properties, axioms and relationships defined into the core or domain ontology.

3 The authorial layers

Any Akoma Ntoso document is based on a legal resource that has been endorsed by an authority empowered by law: an act approved by a parliament, a decree issued by a ministry, a judgement entered by a court. Fidelity to the approved text is, thus, of primary importance; the data layers of Akoma Ntoso have been designed so that it is possible to markup a received legal text while preserving all the information contained in it and changing its content in no way.

The documents that Akoma Ntoso deals with are legal resources whose significance is given by the fact that they have the power to influence citizens' life. Legal texts must, thus, be handled with extreme care and all the measures should be taken to make sure that the technological tools employed to manipulate the texts do not change or interfere with their intended meaning.

In Akoma Ntoso, legal documents are created by enclosing parts of the legal text in XML tags (mixed content model). No pieces of the legal text are discarded, even those that could be generated by an application (e.g. the article numbers in an act). The resulting documents are thus augmented versions of the authentic text; the approved text can be retrieved by simply removing all the XML tags.

The Akoma Ntoso markup process strives to preserve the legal validity of the text as endorsed by the official authority, without adding any additional content to the text. Obviously, the mere act of marking up a sentence involves an act of interpretation or annotation and thus cannot be considered perfectly neutral. However, the kind of markup done at the Akoma Ntoso data layers is almost objective, to the point that some automatic parsers have been developed [11], and is rarely subject to disputes. For this reason Akoma Ntoso documents are designed to contain only one interpretation of the text and structure layers.

In addition to the importance of associating content fragments to their structural roles, the markup in the text and structure layers also provide anchors that the upper layers can

use to give meaning to pieces of text. For instance, to describe that a certain paragraph states a textual modification to a certain act, the relevant text would be marked up with a `mod` element and given an identifier via its `id` attribute. At the same time, in the metadata section, there will be one (or more) `textualMod` elements referring to the URI of the `mod` element that will link that piece of text to its semantic description. The following sections contains examples of how `textualMod` and other legal analysis are connected to the authorial layer.

4 The editorial layer

The Akoma Ntoso metadata layer is a collection of pieces of legal knowledge that can be added onto a legal text by an editorial team as its personal interpretation of the written text, for example the analysis of the reasoning being performed by the judge while writing a judgement or the explicit consequences of the text of an amendment over an act. These pieces of legal knowledge are often subjective may vary across experts. Instead of forcing a single interpretation, Akoma Ntoso allows multiple, and even contrasting, interpretations to be put in the same document. These interpretations of the underlying text form the foundation upon which semantic technologies can make inferences (as discussed in Section 5).

The metadata layer allows agents to provide different kinds of information. The following are examples of the information that can be added with Akoma Ntoso.

Reference disambiguation. The `references` section links pieces of text to ontological entities. The usefulness of this information is twofold. First, conflicts between ambiguous phrases are resolved: for instance, in a sentence of a speech the text "Speaker" may be related to the role called "Speaker" (for example in the sentence "the Speaker must be at least of age 30") while in another sentence the same text "Speaker" may refer to a specific person that is in charge as speaker at the very time the debate was held. In such a scenario, the `references` section will contain these two elements (the `href` attributes point to a URI defined by the Akoma Ntoso naming convention that will be explained in the next section):

```
<TLCRole id="speaker" showAs="Speaker"
href="/ontology/Role/gha/speaker"/>
<TLCPerson id="speaker_20090218" showAs="Speaker"
href="/ontology/Person/gha/parliament/JohnSmith.1948"/>
```

Inside the document, occurrences of the word "Speaker" with the former meaning will be linked to `#speaker`, occurrences of the word "Speaker" with the latter meaning will be linked to `# speaker_20090218`.

In addition to disambiguation, references are used also to consolidate different spellings found in the text to a single entity; in a court judgement, phrases like "Ms. Poliey", "Judge Poliey" and "Her Honour" can all be linked to the same person identified by the TLC Person instance `/ontology/Person/Poliey.1954`.

Legal analysis. The `analysis` section provides information about many legal aspects that can be inferred by a legal expert when interpreting the text. An example is the interpretation of the effects of an amendment in an amendment act. The following example states that, when the act will enter into force, part of the text of the document identified by `/ke/act/1997-08-22/3/main` will be changed: inside the section `#sec34-`

`sub2-itma` of `/ke/act/1997-08-22/3/main`, the text that matches the content of `#mod10-qtd1` (relative to the current document) will be substituted with the content of `#mod10-qtd2` (again, relative to the current document).

```
<textualMod type="substitution" id="am5">
    <source href="#sec4-sub1-itma"/>
    <destination href="/ke/act/1997-08-22/3/main#sec34-sub2-itma"/>
    <old href="#mod10-qtd1"/>
    <new href="#mod10-qtd2"/>
</textualMod>
```

Another example is the identification of the role played by citations of precedents in the judgement argumentation of a judge (e.g., the application of a rule of law of a precedent, the override of a previous ruling, etc.). In the following excerpt it is pointed out that the sentence `#ref01` (a reference to the judgement identified by the URI `/gb/judgement/1829/QB273/eng@/main.xml`) is used to supports the court's decision to deny the request of statutory damages.

```
<judicial>
    <result type="deny"/>
    <supports id="jdc01">
        <source href="#ref01"/>
        <destination href="/gb/judgement/1829/QB273/eng@/main.xml"/>
    </supports>
</judicial>
```

Work identification. The identification section classifies the document using a conceptual model drawn from FRBR (Functional Requirements for Bibliographic Records) [7]. This classification is used to inform the semantic tools that the document is the XML rendering (a *manifestation* in the FRBR model) of a certain version (an *expression* in FRBR) of a document (a *work*), so that we can formally distinguish between different aspects of the idea of document. A detailed account of FRBR in Akoma Ntoso can be found in [20].

In addition to these kinds of metadata, there are other types of metadata currently defined (e.g., *lifecycle* and *workflow* elements for tracking the events affecting the document) and other are being added as Akoma Ntoso extends to its reach to more and more types of analysis of the legal text.

All the information gathered in the metadata layer is derived from the legal text (using the data layers) though subjective reasoning. Many different interpretation can arise over the same legal text from different legal experts. Take, for instance, the following sentence: “the subsection 3 of the section 42 states a modification of the section 44 of the same act”. Two different actors may disagree on the interpretation of that sentence: one sees it as an *authentic interpretation*, another as a *derogation*. From a the legal point of view, the two types of modification produce different effects: the authentic interpretation is applied *ex-tunc* (since the beginning), while the derogation is an exception under some condition. Akoma Ntoso allows both interpretations to coexists in the same document, even if they are in contrast.

5 Semantic technologies and reasoning on Akoma Ntoso documents

Currently, there are interesting developments in the area of legal knowledge representation and manipulation. Akoma Ntoso documents, with their rich metadata layer, can serve as the basis upon which various tools can work on. For example representations expressed at the metadata layer can be used to generate a legal ontology to be used by legal rule modelling technologies like RuleML [6] or the more specialized LKIF [5].

Akoma Ntoso documents are not tied to a particular semantic technology or to particular ontology. The current format is very loose and permits the conversion of information into more specific data models (like RDF [9], OWL [19] or Topic Maps [15]). This strategy warrants that semantic technologies of the future decades will be able to convert Akoma Ntoso documents into their own format without going through what, by then, may be seen as ancient formats or data models.

Even if Akoma Ntoso does not impose which ontology to use to represent legal and legislative concepts, it relies on the FRBR [7] concepts of Work, Manifestation, Expression and Item to describe the relation of a certain document to other documents – for example to distinguish references to the generic Wildlife Act (Work) from references to the Wildlife Act valid in 2008 (Manifestation) and to references to a document that contains the said Wildlife Act valid in 2008 (Expression).

5.1 Representation of facts in Akoma Ntoso

Akoma Ntoso documents can be seen as containers of statements written in some legal human language. The formalization of these statements must take into account many subtle distinctions in order to carry all the meaning that the legal systems pose on these statements.

All statements in an Akoma Ntoso documents follow this schema:

> the **author of a manifestation** (XML rendering) asserts on the **manifestation date** that **the author of the corresponding expression** (version with a specific content) asserts on the **expression date** in a particular **context** that **subject** does **predicate** on **object**.

We can formalize this schema with an octuple:

> (manifestation_author, manifestation_time, expression_author, expression_time, context, subject, predicate, object)

Consider for instance the example already illustrated in the previous sections:

```
<analysis source="#cirsfid">
    <textualMod type="substitution" id="am5">
        <source href="#sec4-sub1-itma"/>
        <destination href="/ke/act/1997-08-22/3/main#sec34-sub2-itma"/>
        <old href="#mod10-qtd1"/>
        <new href="#mod10-qtd2"/>
    </textualMod>
```

```
</analysis>
<identification source="#cirsfid">
    <FRBRExpression>
        <FRBRthis value="/ke/act/2003-12-10/8/eng@"/>
        <FRBRdate date="2003-12-10" name="Generation"/>
        <FRBRauthor href="#parliament" as="#author"/>
    </FRBRExpression>
    <FRBRManifestation>
        <FRBRthis value="/ke/act/2003-12-10/8/eng.akn"/>
        <FRBRdate date="2007-07-27" name="Generation"/>
        <FRBRauthor href="#cirsfid" as="#editor"/>
    </FRBRManifestation>
</identification>
```

One of the assertions contained in the above extract is that

> *#cirsfid* (editor of the manifestation) asserts on 27th July, 2007 (date of the manifestation) that the #parliament (author of the expression) asserts on 10th December, 2003 (date of the expression) that in the context of the fragment with id *#art4-cla1-itma* the text modification *am5* has been performed which is of type substitution

or, using an octuple

> (#cirsfid, “2007-07-27”, #parliament, “2003-12-10”, #art4-cla1-itma, #am5, isOfType, “substitution”)

It must be pointed out that not all the parts of these assertions lay at the same FRBR level: some are inherently properties of the FRBR manifestation (the date when `#cirsfid` did add its own metadata to the Akoma Ntoso file containing the act) while other are properties of the connected FRBR expression (the date when `#parliament` did approve the act). For this reason, every Akoma Ntoso document explicitly states what are the FRBR manifestation and expression that it is modelling inside the `FRBRthis` element. The `value` attribute of `FRBRthis` specifies the URI of the document on which all the assertions are made.

5.2 Everything is semantically generic

Akoma Ntoso defines an ontological structure for metadata that is grounded on what we call the *Top Level Classes* (*TLC*). They do not define a real ontology: on purpose, none of the TLCs has a precise meaning nor a well-defined set of properties. The only constraint imposed on them is that all the instances of a certain TLC must follow the Akoma Ntoso naming guideline in the definition of their URI [17]. This choice is driven by a precise design intent: the ontology used to manage the concepts expressed in Akoma Ntoso documents is not hard coded in the document, but can be decided at will by individual users and applications, according to their particular needs.

Furthermore, TLCs, even if represented by a clear label and a specific URI, have neither formal (logically defined) nor informal (written in natural language) semantics. Akoma Ntoso does not define the classes of a particular ontology but only the URIs and labels that should be used for the classes defined in all ontologies.

This assumption is needed to allow a great degree of flexibility in what can be expressed in the metadata layer of Akoma Ntoso documents, in order to adapt any

legal document to any ontological representation of concepts. It is the duty of a third party, possibly the document editor, possibly its users, to associate a clear and formal semantics to each class using a specific formalism (e.g. OWL). This semantic genericness, coupled with XML elements whose names resemble familiar concepts, is an important feature that allows Akoma Ntoso to maintain documents understandable and consumable independently from the passing of time: future toolmakers ("The 'future toolmaker' is 10 years old now." [18]) will have clues about the intended meaning of a marker even in the unfortunate case the formal ontology is no longer available.

Akoma Ntoso makes ten different and disjoint TLCs available to document creators in order to identify individual entities present in the document: Person, Organization, Concept, Object, Event, Place, Process, Role, Term, Reference (all identified by the URIs `/ontology/[ TLC ]`).

Using these TLCs and the canonical FRBR classes Work, Expression, Manifestation and Item, it is possible to make quite complex assertions on Akoma Ntoso documents. Akoma Ntoso does not aim at describing neither objective facts nor personal opinions about such classes according to the author of the document: rather it allows to express an interpretation that is due, in a precise moment, to a particular actor working on the statements that can be found in the published legal text (e.g., the interpretation of `#cirsfid` of the correct interpretation of the nature of the textual modification `am5`).

By defining a mechanism for describing items (actors, legal documents, properties, concepts, etc.) that are involved in the assertions (both making assertions and being the subject of an assertion), and by relating them to some TLC, we are able to assert facts about an Akoma Ntoso document, implicitly producing a data model for its semantic data.

5.3 Ontology URI naming conversion

All the items in an Akoma Ntoso document can belong to a particular TLC simply by specifying an URI that follows a naming convention [17]. The following example shows a list of URIs pointing to the same entity.

```
/ontology/Person/lewanika.1961
/ontology/Person/judges/lewanika.1961
/ontology/Person/za/judges/lewanika.1961
```

Even if a human can interpret them in some ways depending on the particular interpretation used, from an ontological point of view all these three URIs say the same thing: they identify an instance of the top level class Person and they identify exactly the same instance in any ontology (regardless of the ontology).

The Akoma Ntoso naming convention contains few precise rules:

- the first two fragments of the URI (`/ontology/Person`) specify the TLC the instance belongs to;
- the last fragment of the URI (`lewanika.1961` in the above example) is the identifier of the instance we are referring to (a South African judge called Lewanika and born in 1961);
- the middle URI fragments (`za/judges/`), when they are present, provide evocative information for the human reader and for the systems that cannot

use any more detailed ontology for the document. They suggest, in fact, a clear interpretation from a human perspective, e.g., that Lewanika is a South African person holding the role of judge. Since Akoma Ntoso does not force any given set of properties for the top level classes, the responsibility to choose which, and whether, additional fragments should be added lies with the author of the manifestation.

Taking into account the implicit semantics given to each URI by the Akoma Ntoso naming convention, it is possible to query an Akoma Ntoso-compliant legal XML document using XPath [2], without relying on external ontologies. The following excerpt shows some references to various resource.

```
<akomaNtoso xmlns="http://www.akomantoso.org/1.0">
    <references source="#cirsfid">
        <TLCOrganization id="parliament"
            showAs="Parliament of Kenya"
            href="/ontology/Organization/ke/parliament" />
        <TLCOrganization id="cirsfid" showAs="CIRSFID"
            href="/ontology/Organization/cirsfid" />
        <TLCPerson id="fv" showAs="Fabio Vitali"
            href="/ontology/Person/fv" />
        <TLCRole id="author" showAs="Author"
            href="/ontology/Roles/author" />
        <TLCRole id="jurist" showAs="Jurist"
            href="/ontology/Roles/jurist" />
        <TLCRole id="editor" showAs="Editor"
            href="/ontology/Roles/editor" />
    </references>
    [...]
```

Through XPath, it is possible to perform queries based on the semantic data that is present in the document. For example, the following queries could be performed on the data shown in the above excerpt:

- what roles have been involved in the generation of the legal document: `//references/element()[matches(@href,'/Roles/')]`;
- which Kenyan organizations are referred to in the legal document: `//references/element()[matches(@href,'/Organization/.*/ke/')]`.

The naming convention used by Akoma Ntoso, together with the presence of additional middle URI fragments, allows the semantic data available in the document to be queried, even in sophisticated ways, without requiring access to any ontology. When the underlying ontology is available, the additional information provided by the middle URI fragments can be discarded and more complex queries, based on deductive elements such as those gathered through the use of reasoners, can be performed. The optional middle URI fragments fulfil, thus, two different tasks: on the one hand they provide clues to the human readers who do not have access to the underlying ontologies, on the other hand they allow semantic data about the referenced entities to be carried also in the document itself and not only in external knowledge bases.

5.4 Transformation into a proper semantic data model format

In order to be able to carry out more complex queries on the facts expressed in Akoma Ntoso documents, it is necessary to bind the abstract Akoma Ntoso classes and instances to external ontological models. This also require to translate the same facts to a more concrete representations that the current semantic tools can work with (e.g., RDF and Topic Maps). This happens without specifying a particular format, but allowing users to choose their favourite formats and tools, a choice that the users will take on the base of current languages, their knowledge and other technical constraints.

An example of a concrete data model format that today fits the current technology scenario is RDF [9]. Although the following example, covering the extraction of metadata from Akoma Ntoso documents, is completely based on RDF, it is also possible to use different end-format for expressing these semantic data, such as Topic Maps [15]. In any case, if we wanted to convert an XML document into a set of RDF/ Topic Maps assertions, we could use a GRDDL transformation. GRDDL (Gleaning Resource Descriptions from Dialects of Languages) [4] is a W3C Recommendation that standardize the extraction of semantic data from XML documents using one or more XSLT stylesheets [8], obtaining, in the particular example presented, an RDF document. Note that this particular mechanism it is also suggested in the current CEN Metalex proposal [3] of which Akoma Ntoso is a compliant instance:

> If metadata is not available as RDFa, it must be systematically translatable from the custom format to RDF. The translation from a proprietary metadata format to RDF must be publicly available following the Gleaning Resource Descriptions from Dialects of Languages (GRDDL) specification.

It is easy to convert an Akoma Ntoso document in one of the semantic formats currently available. For example, if we consider the FRBR class implicitly defined within the Akoma Ntoso schema, we could generate assertions that link FRBR instances gleaned from a document to the OWL-specified FRBR ontology, simply through GRDDL. Given the following fragment:

```
<identification source="#cirsfid">
    <FRBRExpression>
        <FRBRthis value="/ke/act/2003-12-10/8/eng@"/>
        <FRBRdate date="2003-12-10" name="Generation"/>
        <FRBRauthor href="#parliament" as="#author"/>
    </FRBRExpression>
    <FRBRManifestation>
        <FRBRthis value="/ke/act/2003-12-10/8/eng.akn"/>
        <FRBRdate date="2007-07-27" name="Generation" />
        <FRBRauthor href="#cirsfid" as="#editor" />
    </FRBRManifestation>
</identification>

<references source="#cirsfid">
    <TLCOrganization id="parliament" showAs="Parliament of Kenya"
     href="/ontology/Organization/ke/parliament"/>
    <TLCRole id="author" showAs="Author" href="/ontology/Role/author"/>
</references>
```

we can use GRDDL to obtain assertions such as the following:

```
@prefix ke: <http://www.akomantoso.org/ke> .
@prefix an: <http://www.akomantoso.org/ontology/> .

ke:/act/2003-12-10/8/eng@ a an:FRBRExpression ;
    an:hasDate "2003-12-10" ;
    an:hasAuthor an:Organization/ke/parliament .
an:Organization/ke/parliament an:hasRole an:Role/author .
```

These are just plain RDF statements unrelated to any logical structure. To address particular demands (such as reasoning, data sharing and so on) we need to associate those instances, classes and properties to well-defined ontologies. Though an additional XSLT stylesheet in the GRDDL process we can add new ontological data about FRBR:

```
@prefix frbr: <http://purl.org/vocab/frbr/core#> .
@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .

an:FRBRExpression owl:sameAs frbr:Expression .
ke:/act/2003-12-10/8/eng@ frbr:realizer an:person/Organization/ke/
parliament .
an:hasDate owl:sameAs dc:date .
```

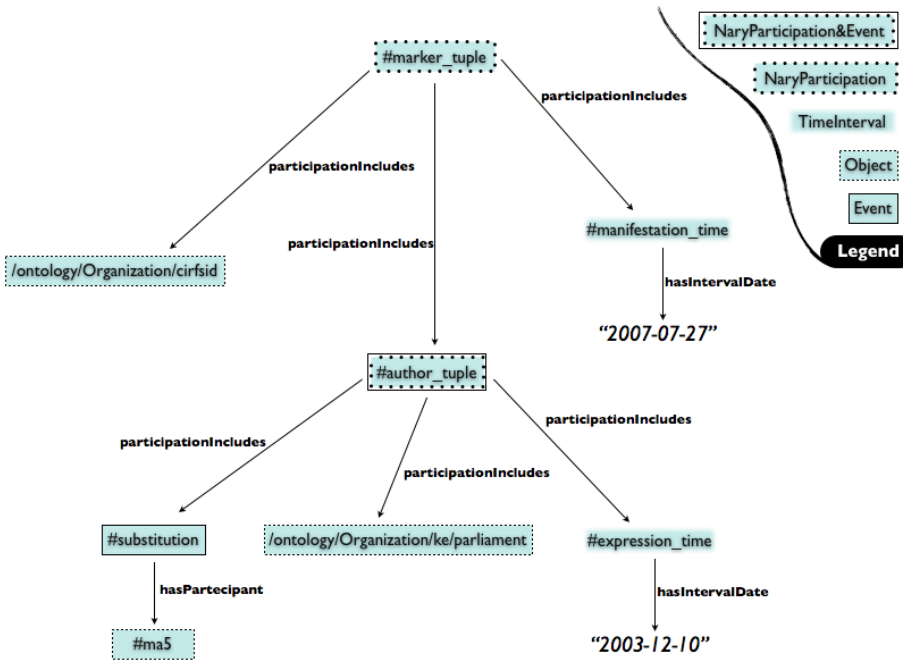

Fig. 1. A possible Akoma Ntoso data extraction into a OWL pattern-based format of the octuple introduced in Section 5.4.

Unfortunately, even if we can define semantics in order to infer automatically new data, OWL does not have a native or standard way to model reified statements. To represent in OWL the Akoma Ntoso octuple statement illustrated previously, we

have to resort to peculiar approaches. The NeOn project [14] has developed many elegant and logically-sound techniques to overcome the limitations of OWL; they have been collected under the name of Ontology Design Pattern [13]. The Ontology Design Patterns are a set of guidelines that help the design process of ontologies; each pattern codifies one or more best practices of the ontology-design realm. The following examples will use and briefly introduce some of those patterns; a proper description of each used pattern is out of the scope of our paper.

The most straightforward way to handle complex assertions, such as the octuples used by Akoma Ntoso, in OWL is to use the *n-ary participation* pattern. The n-ary participation pattern is used to describe events happening in a certain moment and that involve one or more entities. This pattern can also be used to simulate RDF reifications in OWL. Fig. 1 shows a graphical representation of the previously shown octuple expressed in OWL using the n-ary participation pattern.

5.5 Querying Akoma Ntoso documents

Having an RDF/OWL or a Topic Maps representation of Akoma Ntoso data could be very useful in querying metadata of a legislative document using different and much more proper tools than XPath to query data models, such as SPARQL [1] and TMQL [16] according to the format we use to express Akoma Ntoso metadata. Let us take into consideration the following excerpt from an Akoma Ntoso document:

```
<identification source="#cirsfid">
    <FRBRWork>
        <FRBRthis value="/ke/act/2003-12-10/8"/>
        <FRBRauthor href="#parliament" as="#author"/>
    </FRBRWork>
    <FRBRExpression>
        <FRBRthis value="/ke/act/2003-12-10/8/eng@"/>
        <FRBRauthor href="#parliament" as="#author"/>
    </FRBRExpression>
    <FRBRManifestation>
        <FRBRauthor href="#cirsfid" as="#editor"/>
        <FRBRauthor href="#fv" as="#editor"/>
    </FRBRManifestation>
</identification>

<references source="#cirsfid">
    <TLCOrganization id="cirsfid" href="/ontology/Organization/cirsfid"/
>
    <TLCPerson id="fv" href="/ontology/Person/fv" />
</references>
```

Suppose we want to query it looking for whether the document is a candidate to be an authoritative Expression of a given Work (i.e., the author of the Expression is the same of the Work).

We can describe this request in a query comparing two different XPath expressions:

```
//FRBRExpression/FRBRauthor/@href = //FRBRWork/FRBRauthor/@href
```

Obviously, querying the Akoma Ntoso metadata through XPath is possible, but it can be also intricate: XPath was developed to browse easily XML hierarchies, that are trees, not to query much more complicate data structures (i.e., a graph) such as the one

implicitly defined through Akoma Ntoso metadata. The previous XPath expressions give the correct answers for the respective queries but the way in which a user query the document is not so natural.

Converting the Akoma Ntoso metadata into a data model format, such as RDF or Topic Maps, can help to build queries in a much easier way, using specific languages (e.g., SPARQL or TMQL) expressly developed to handle these scenarios. Let us take into consideration an OWL ontology[1] describing the FRBR data shown in the previous excerpt. Then, we can query the ontology using a SPARQL representation of the previous XPath expressions:

```
ASK {
    ?e frbr:realizer ?a .
    ?w frbr:creator ?a .
    ?e akomantoso:frbrthis "/ke/act/2003-12-10/8/eng@" .
    ?e frbr:realizes ?w .
}
```

The conversion from an Akoma Ntoso document into an OWL document can be done in an automatic way using GRDDL [4]. Moreover, this transformation is completely free of constraints: we can use whatever ontologies we prefer to represent those data according to some particular model-representation needs.

In order to demonstrate the feasibility of the GRDDL approach, we developed an XSLT that extracts metadata from an Akoma Ntoso document according to three different ontologies – the FRBR OWL ontology, the n-ary participation pattern ontology, and an ontology for handling other data concerning textual modifications.

Let us give an example: the XSLT template to transform the FRBR manifestation of an Akoma Ntoso document into RDF statements:

```
<xsl:template match="an:FRBRManifestation">
    <frbr:Manifestation
        rdf:about="{//an:FRBRManifestation/an:FRBRthis/@value}">
        <xsl:for-each select="an:FRBRauthor">
            <frbr:producer rdf:resource="{$uri}
                {//an:references/element()
                    [@id = substring-after(@href,'#')]/@href}" />
        </xsl:for-each>
        <dc:date><xsl:value-of select="an:FRBRdate/@date" /></dc:date>
    </frbr:Manifestation>
</xsl:template>
```

Using this template, we are able to map any element and attribute involved in the description of an FRBR manifestation of any Akoma Ntoso document onto proper classes and individuals, in particular:

- Akoma Ntoso *FRBRManifestation* elements are represented by the class *Manifestation* of the FRBR ontology;
- all the URIs contained in the attribute *value* of the element *FRBRthis* of any *FRBRManifestation* become proper individuals belonging to the class *Manifestation*;

[1] http://purl.org/vocab/frbr/core.

- each *FRBRauthor* contained in a particular *FRBRManifestation* is mapped adding an RDF statement having as subject the particular ontological individual of that manifestation, as object the resource identified by the URI specified as value of the attribute *id* of a TLC, obtained by dereferencing the attribute *href* of the element *FRBRauthor*, and, as predicate, the object property *producer* of the FRBR ontology;
- the value of the attribute *date* of the element *FRBRdate* contained in *FRBRManifestation* becomes the object literal of the RDF statement obtained using, as subject, the ontological individual refers to the manifestation we are processing, and, as predicate, the Dublin Core [10] property *date*.

Applying this template on an Akoma Ntoso document, we obtain the following RDF statements:

```
@prefix m: <http://www.akomantoso.org/> .

m:ke/act/2003-12-10/8/eng.akn a frbr:Manifestation ;
    dc:date "2007-07-27" ;
    frbr:producer m:ontology/Person/fv , m:ontology/Organization/
cirsfid .
```

Similarly, we wrote a template based on the n-ary participation pattern and our textual modifications model that, for instance, allows us to export the relative Akoma Ntoso metadata of the excerpt shown in Section 5.1 in the following statements:

```
m:manifestation_time a ti:TimeInterval ;
    ti:hasIntervalDate "2007-07-27" .

m:ontology/Organization/ke/parliament a part:Object .

m:author_tuple_5 a nary:NaryParticipation ;
    nary:participationIncludes m:ontology/Organization/ke/parliament ,
        m:author_tuple_5_tmod , m:expression_time .

m:author_tuple_5_tmod a m:substitution ;
    part:hasParticipant m:ke/act/2003-12-10/8/eng.akn#am5 .

m:marker_tuple_5 a nary:NaryParticipation ;
    nary:participationIncludes m:manifestation_time ,
        m:author_tuple_5 , m:ontology/Person/fv ,
        m:ontology/Organization/cirsfid .

m:substitution a owl:Class ; rdfs:subClassOf part:Event .

m:ontology/Person/fv a part:Object .

m:ke/act/2003-12-10/8/eng.akn#am5 a tm:TextualModification ;
    tm:hasDestination m:ke/act/1997-08-22/3/eng/main#art34-cla2-itma ;
tm:hasNewText "statement of assets and liabilities" ;
tm:hasOldText "balance sheet" ;
tm:hasSource m:ke/act/1997-08-22/3/eng/main#art34-cla2-itma ,
        m:ke/act/2003-12-10/8/eng.akn#art4-cla1-itma .

m:expression_time a ti:TimeInterval ; ti:hasIntervalDate "2003-12-10" .

m:ontology/Organization/cirsfid a part:Object .
```

These pieces of code show that the creation of a GRDDL process to extract metadata from an Akoma Ntoso document is quite feasible and moreover is also very flexible and adaptable according to particular needs. Apart from being able to query data through SPARQL or TMQL, modelling Akoma Ntoso metadata by OWL or Topic Maps it is possible to use reasoners to infer new assertions on the basis of all the extracted data from Akoma Ntoso documents available.

6 Conclusions

Akoma Ntoso has been designed as a format for legal documents that must be read and understood for decades and at the same time be useful to computer applications, including semantic reasoners. In order to balance clearness, fidelity to the authentic legal text, interoperability and usability with semantic tools, Akoma Ntoso made some clear design choices. In this paper we showed how these choices fit the stated goals: using XML as the underlying mark-up format and having clearly separated layers allow documents to be preserved for long periods of time and without modifications to the endorsed texts. Additionally, multiple agents can provide their own interpretation of certain legal aspects of the given legal text. Moreover, computer reasoners can extract semantic information from Akoma Ntoso documents and reason over them both with or without user-supplied ontologies.

The approach used by Akoma Ntoso allows the development of systems that use more sophisticated formal logic modelling framework, like non-monotonic or non-deductive logics in order to apply sophisticated legal reasoning theories, more suitable for the complex legal domain, filling the gap between all the semantic web layers while preserving interdependency and expressiveness.

Acknowledgments. The authors would like to thank Flavio Zeni (Kenya Unit of UN/DESA) and all of his team, for supporting the Akoma Ntoso project and for permitting this research project. We would also like to thank Alexander Boer and Enrico Francesconi for their precious contributions on the definition of the URI syntax.

References

[1] Beckett D., Broekstra J. (2008). SPARQL Query Language for RDF. W3C Recommendation. World Wide Web Consortium. http://www.w3.org/TR/rdf-sparql-query/.

[2] Berglund, A., Boag, S., Chamberlin, D., Fernández, M. F., Kay, M., Robie, J., Siméon, J. (2007). XML Path Language (XPath) 2.0. W3C Recommendation. World Wide Web Consortium. http://www.w3.org/TR/xpath20/.

[3] Boer, A., Vitali, F., Palmirani, M., Retai, B. (2009). CEN Metalex workshop agreement. http://www.metalex.eu/WA/proposal.

[4] Connolly, D. (2007). Gleaning Resource Descriptions from Dialects of Languages (GRDDL). W3C Recommendation. World Wide Web Consortium. http://www.w3.org/TR/grddl/

[5] Gordon, T.F. (2008. Constructing Legal Arguments with Rules in the Legal Knowledge Interchange Format (LKIF). In Computable Models of the Law: Languages, Dialogues, Games, Ontologies, pp. 162-184.

[6] Hirtle, D., Boley, H., Grosof, B., Kifer, M., Sintek, M., Tabet, S., Wagner, G. (2006). Schema Specification of RuleML 0.91. http://ruleml.org/0.91/.

[7] IFLA Study Group on the Functional Requirements for Bibliographic Records. (1998). Functional requirements for bibliographic records: final report. K. G. Saur.

[8] Kay, M (2007). XSL Transformations (XSLT) Version 2.0. W3C Recommendation. World Wide Web Consortium. http://www.w3.org/TR/xslt20.

[9] Manola, F., Miller, E. (2004). RDF Primer. W3C Recommendation. World Wide Web Consortium. http://www.w3.org/TR/rdf-primer/.

[10] Nilsson, M., Powell, A., Johnston, P., Naeve, A. (2007). Expressing Dublin Core metadata using the Resource Description Framework (RDF). http:// dublincore.org//documents/2007/06/04/dc-rdf.

[11] Palmirani, M., Benigni, F. (2007). Norma-System: A Legal Information System for Managing Time. In V Legislative XML Workshop, pp. 205-224.

[12] Palmirani, M., Contissa, G., Rubino, R. (2009). Fill the gap in the legal knowledge modelling. In the Proceeding of RuleML 2009, pp. 305-314.

[13] Presutti, V., Gangemi, A. (2008). Content ontology design patterns as practical building blocks for web ontologies. In the Proceedings of 27th International Conference on Conceptual Modeling (ER2008). Barcelona, Spain.

[14] Presutti, V., Gangemi, A., David, S., Aguado de Cea, G., Suárez-Figueroa, M. C., Montiel-Ponsoda, E., Poveda, M. (2008). A Library of Ontology Design Patterns. NeOn project deliverable D2.5.1.

[15] SC34/WG3. (2003). Topic Maps. ISO 13250. International Organization for Standardization. Geneva, Switzerland.

[16] SC34/WG3. (2008). Topic Maps Query Language (TMQL). ISO 18048. International Organization for Standardization. Geneva, Switzerland.

[17] United Nations. (2009). Akoma Ntoso: Naming Conventions. http:// www.akomantoso.org/release-notes/akoma-ntoso-1.0-schema/naming-conventions-1.

[18] United Nations. (2009). Akoma Ntoso: Potential users, http:// www.akomantoso.org/akoma-ntoso-in-detail/users.

[19] W3C OWL Working Group (2009). OWL 2 Web Ontology Language Document Overview. W3C Recommendation. World Wide Web Consortium. http:// www.w3.org/TR/owl2-overview/.

[20] de Oliveira Lima, A., Palmirani, M., Vitali, F. (2008). Moving in the Time: An Ontology for Identifying Legal Resources. In Computable Models of the Law, Languages, Dialogues, Games, Ontologies pp. 71-85.