Towards a country-independent data format: the Akoma Ntoso experience

Fabio Vitali*, Flavio Zeni°

*Department of Computer Science, University of Bologna, fabio@cs.unibo.it
°UN Department of Economic and Social Affairs, Nairobi, Kenya, zeni@un.org

Abstract. AKOMA NTOSO (Architecture for Knowledge-Oriented Management of African Normative Texts using Open Standards and Ontologies) is an operating framework and a set of guidelines for driving e-Parliament services in a Pan-African context by formalizing and harmonizing the storage, publication and exchange of Parliamentary documents using a precise, common and easy to understand data format based on XML. The AKOMA NTOSO XML document schema provides sophisticated description possibilities for Parliamentary document types (including legislative documents and parliamentary records), and supports document structures by systematically relying on international standards, best practices, guidelines and widely recognizable design patterns.

Keywords: AKOMA NTOSO, Open Access, Pan-African Interoperability, Parliamentary records, XML

1. Introduction

In 2004 and 2005, the UNITED NATIONS Department for Economics and Social Affairs (UN/DESA) project, “Strengthening Parliaments’ Information Systems in Africa”, has aimed at empowering legislatures to better fulfil their democratic functions, using ICTs to increase the quality of parliamentary services, facilitate the work of parliamentarians and create new ways to promote the access of civil society to parliamentary processes.

A strategic role in this project is played by the AKOMA NTOSO (Architecture for Knowledge-Oriented Management of African Normative Texts using Open Standards and Ontologies) framework, a set of guidelines for e-Parliament services in a Pan-African context. The framework addresses information content and recommends technical policies and specifications for building and connecting Parliament information systems across Africa.

In particular, the AKOMA NTOSO framework proposes an XML document schema providing sophisticated description possibilities for several Parliamentary document types (including bills, acts and parliamentary records, etc.), therefore fostering easier implementation of Parliamentary Information systems and interoperability across African Parliaments, ultimately allowing open access to Parliamentary information.
The AKOMA NTOSO Framework reaches three main objectives which are instrumental for the success of the overall project:

- to define a common standard for data interchange between parliaments;
- to define the specifications for a base document model on which parliamentary systems can be built;
- to define an easy mechanism for citation and cross referencing of data between parliaments.

The AKOMA NTOSO framework aims at providing two basic types of interoperability: *semantic interoperability* is concerned with ensuring that the precise meaning of exchanged information is understandable by any person or application receiving the data; *technical interoperability* is aimed at ensuring that all AKOMA NTOSO-related applications, systems, interfaces are based on a shared core of technologies, languages and technical assumptions easing data interchange, data access and reuse of acquired competencies and tools. AKOMA NTOSO ensures technical interoperability by enforcing the use of open standards and open document formats, based on the XML (eXtensible Markup Language) language whose specifications are a world-wide standard and for which numerous tools and applications have been developed and are widely available.

In this paper we plan to describe the genesis and the fundamental aspects of the AKOMA NTOSO documents.

2. The need for a country-independent data format

Parliaments are currently exploiting ICT to improve the quality of their services and to improve access to all Parliamentary information: In Africa, this is particularly considered a strategic resource for its young and active democracies. In a Pan African context, Parliaments are also promoting collaboration among Parliaments and co-operation among countries in order to tackle the enormous problems that Africa is facing. At present, most interactions within and among Parliaments require numerous disparate transactions across multiple departments and Parliaments and there is very limited consolidation and aggregation across national Parliaments’ boundaries.

Yet, Parliaments in Africa are continually confronted with demands for ever-greater dialogue between the electors and the elected, and have to examine how the management of information and official documents
can improves transparency and citizens’ access to political decisions, thereby permitting greater understanding of the democratic process. Needless to say, improved access to documents regarding activities of the Parliaments enable citizens to hold Parliaments accountable, stimulate greater efficiency and enhance democracy.

Connecting Parliaments has many benefits: in addition to its value as a knowledge transfer mechanism – whereby one Parliament can learn from the other – it also can be a tremendous boost to Parliamentary positive imitation. By seeing what others are doing, members of Parliaments can discover the possibility of doing the same in their own Parliaments. The explosion of Internet-based systems has increased the possibilities and range of such dialogue but this can be achieved and exploited only if common standards to produce, classify and share Parliamentary and legislative electronic documents are agreed and used by African Parliaments.

Providing access to primary legal materials and parliamentary documents is not just a matter of providing physical or on-line access to them, what we could be termed formal access. What we call open access, on the other hand, requires that the information content and the tools for search and retrieval of data are organized so as to allow users (MPs, the Executive, public administration, enterprises and citizens) to access and use the information in the form that is most convenient to them. In order to build an information system that provides open access, there is the need to establish both common specifications for document models on which parliamentary systems can be built and a mechanism for citation and cross referencing of legal documents. To embark on the development of an information system without defining and agreeing on a standardised way for Parliaments to classify and structure their documents, e.g. bills, debate records, etc., would be to ignore the lessons learnt and the best practices of ICT development.

AKOMA NTOSO1 (Architecture for Knowledge-Oriented Management of African Normative Texts using Open Standards and Ontologies) is an enabling framework for the effective access to and exchange of machine-readable Parliamentary Documents such as legislation, debate record etc. It aims to standardize simple, technology-neutral representations of Parliamentary Documents in order to improve inter-Parliamentary cooperation, and reduce the costs of Parliamentary IT support systems.

1 “Akoma Ntoso” (linked hearts) is the symbol used by the Akan people of West Africa to represent understanding and agreement. Likewise, AKOMA NTOSO represents common standards that provide open access to parliamentary documentation and allow Parliaments to exchange information more efficiently, like “linked hearts” [Akoma Ntoso, in “West African Wisdom: Adinkra Symbols & Meanings”, http://www.welltempered.net/adinkra/htmls/adinkra/akon.htm].
To obtain interoperability between Parliamentary information systems for the purpose of providing open access, it was necessary to develop and adopt a framework for interoperability between IT systems at all levels that could foster co-operation between different institutions, administrations and Parliaments.

The framework takes the form of a collection of specifications aimed at facilitating the data interchange of Parliaments’ systems and services using common open standards that avoid vendor lock-in and allow for greater public access to information.

By bringing together the relevant specifications under an overall framework, IT management and developers can have a single point of reference for the interoperability specifications that should be followed. By adopting these specifications, Parliamentary system designers can ensure interoperability between systems while at the same time enjoying the flexibility of selecting different hardware, operating systems and application software to implement solutions.

In 2004 and 2005 the UNITED NATIONS Department for Economics and Social Affairs (UN/DESA) “Strengthening Parliaments’ Information Systems in Africa” project started to address the need that Parliaments have to exchange information amongst themselves. With the establishment of the Pan African Parliament, this need has become an institutional requirement for the Parliaments that are represented there. It is clear that it would have been inappropriate to proceed with the development and implementation of the Parliamentary Information systems without establishing common guidelines on the structure and classification of digital parliamentary documentation, and it would not be following international best practices in such matters. In practice, it could jeopardize the long-term sustainability of such complex systems.

To address this issue, the Project developed the AKOMA NTOSO framework proposal³, based on the best practices and drawing from the experience of other continents. It was apparent that it would have been imprudent to deploy information systems in the National Parliaments without a Pan-African framework, and relying on no set of recommendations and guidelines for e-Parliament services in a pan-African context.

By adopting AKOMA NTOSO, Parliamentary System designers can ensure interoperability between systems while at the same time enjoy the flexibility of selecting different hardware, systems, and application software to implement specific solutions.

---

² UNDESA, the United Nation Department of Economic and Social Affairs, http://www.un.org/esa/desa/
In order to enlist the support of the African Parliaments, UN/DESA organised, on 10 and 11 February 2005, an International Conference, “Parliaments’ Information Management in Africa: Challenges and Opportunities of ICTs to Strengthen Democracy and Parliamentary Governance”. The participating Parliaments were introduced to the AKOMA NTOSO framework proposal as well as to “Eurovoc-Africa” - a multilingual parliamentary thesaurus; and the proposal for the development of a common, but customisable Parliamentary Information System that included cooperative information services and repositories.

The Conference was attended by speakers and high-level delegations from the Parliaments of Algeria, Angola, Cameroon, Ghana, Kenya, Madagascar, Malawi, Mauritius, Mozambique, Rwanda, Sao Tomé and Príncipe, Somalia, Sudan, Tanzania, Tunisia, Uganda, Zambia and Zimbabwe, the Pan African Parliament, East Africa Legislative Assembly, ECOWAS Parliament, SADC Parliamentary Forum and the Italian Chamber of Deputies. Parliamentary officials from the above countries as well as from South Africa, the European Parliament, some European Union countries, representatives from international organisations, development partners, the private sector, and civil society organisations also attended the Conference.

The Conference unanimously adopted the “Nairobi Declaration” which, among other issues, recommends to promote the exchange of information by supporting the AKOMA NTOSO framework and a common Parliamentary multilingual thesaurus thereby establishing “a coordinating mechanism to adopt, promote and maintain the Pan-African Parliamentary Interoperability (PAPI) proposal” and through the development of Parliamentary Information Systems “based on open and interoperable standards, innovative technologies and full access to information”.

AKOMA NTOSO was developed as a necessary foundation for the development of a comprehensive Parliamentary Information System (PIS). The goal of the Parliamentary Information System is to maximize the operational efficiency and effectiveness of National Legislatures by implementing a solution which provides secure, reliable, and timely collection, storage, access, and transmission of information. The system will equip Parliaments with a solution that fosters accessibility, transparency and accountability of Parliaments by exploiting open source multi-platform applications based on open standards and available in multiple human languages.

PAPI was later renamed to “AKOMA NTOSO”. This was done in order to give the framework, as requested in the Nairobi Conference, an African name to reflect its Pan African nature and ownership.
Individual country Parliaments should use the guidance provided by the AKOMA NTOSO Framework to supplement their national e-Government initiatives with a Pan-African dimension and thus enable Pan-African interoperability of Parliaments. Thus the AKOMA NTOSO framework is meant to supplement, rather than replace, national interoperability guidance that may exist, and to add a pan-African dimension to them.

3. Objectives

AKOMA NTOSO (Vitali and Zeni, 2006) is an enabling framework for the effective exchange of machine readable Parliamentary Documents such as legislation, debate record etc. AKOMA NTOSO aims to standardize simple, technology-neutral representations of Parliamentary Documents in order to improve inter-Parliamentary cooperation, and reduce the costs of Parliamentary IT support systems. Parliaments are major producers of data and information that are vital for the democratic well-being of a country and the lifeblood of political participation. The lack of a standardized way for Parliaments to classify and structure their data resources, information technology and business processes stands in the way of increased integration of information exchange and this in turn limits the efficiency and effectiveness of Parliamentary activities.

The goal of AKOMA NTOSO is to fulfil the citizens’ right to access Parliamentary proceedings and deliberations and to support Parliaments in managing legislative documentation life-cycles efficiently. This is facilitated through the definition and adoption of technical rules to improve data exchange, document life-cycle automation and standardized representations of data and metadata in the African Parliamentary domain. This goal can be further specified in three major subgoals:

3.1. Data interchange between parliaments

Parliaments function through the medium of documents. Debate in Parliamentary chambers is recorded as documents. Legislation is passed through the voting process via a combination of documents – the proposed legislation itself, proposed amendments, committee working papers and so on.

Given that the process is document-centric, the key enabler of streamlined Information Technology in Parliaments is the use of open document formats for the principal types of documents. This allows easy exchange and aggregation of Parliamentary information – in addition
to reducing the time required to make the information accessible via different electronic publishing media.

The Information Technology industry has coalesced around a standard technology for open document formats known as XML (eXtensible Markup Language)\(^5\). The AKOMA NTOSO framework uses of industry standard XML to define a comprehensive set of XML-based Parliamentary document formats. This comprehensive set open document formats includes:

- Primary Legislation – covering the lifecycle of a piece of legislation
- Parliamentary Debates
- Amendment lists
- Committee briefs
- Journals

3.2. A schema for African parliamentary metadata

Metadata is structured information about a resource. Metadata enables a resource to be found by indicating what the resource is about and how it can be accessed with a series of structured descriptions. Metadata facilitates the discovery and use of online resources by providing information that aids and increases the ease with which information can be located by search engines that index metadata.

The AKOMA NTOSO metadata format is primarily concerned with resource discovery and records management. The aim of this metadata format is to ensure that people searching the Parliaments information space online have fast and efficient access to descriptions of many different resources. The access to XML documents thanks to such shared metadata allows the creation of advanced search and retrieval functions across heterogeneous databases.

Structured collections of shared metadata will generate a thesaurus of legislative terms even across disparate documents, and documents occurring in disparate locations. These in turns will allow faster and more precise searches and categorizations of available legislative documents. The AKOMA NTOSO metadata format is designed to be extensible, so that those Parliaments with different, or more specific, metadata needs may add those extra elements and qualifiers needed to meet their own requirements.

\(^5\) XML, the extensible Markup Language, \url{http://www.w3.org/XML}
3.3. Easy citation and referencing

Official documents, bills, laws and acts contain numerous references to other official documents, bills, laws and acts. The adoption of a common naming convention to reference a distributed document corpus like the one of the African Parliaments greatly enhances the accessibility and richness of cross references.

The naming convention enables comprehensive cross referencing and hyper-linking, so vital to any Parliamentary corpus, for instance:

− From debate records into legislation;
− From sections of legislation to sections of legislation in the same act;
− From sections of legislation to sections of legislation in another act of the same Parliament;
− From sections of legislation to sections of legislation in another act of the same Parliament;
− From sections of legislation to sections of legislation of a different institution, such as acts of lesser authorities, acts of different countries, international treaties, or acts of superior authorities such as the Pan African Parliament.

The AKOMA NTOSO Naming Convention is intended to enable a persistent, location-independent, resource identification mechanism. The adoption of a scheme based on this Naming Convention allows the full automation of distributed hypertext.

4. Key issues in the development of AKOMA NTOSO

Developing a full model for legislative documents requires full awareness of a particularly complex environment of uses and applications and an environment that widely extends in space, time, and availability.

In particular, thinking about long term storage and an open set of applications requires making format choices completely independent of the tools that will be used for them, and ensuring that new tools can be used now and in an unspecified future for tasks as yet unclear or completely unknown.

In many ways, the AKOMA NTOSO data model derives from an analogous initiative brought forth in Italy by the National Centre for Informatics in Public Administration (Centro Nazionale per l’Informatica nella Pubblica Amministrazione or CNIPA\(^6\)), a project called NormeIn-

\(^6\) CNIPA, Centro Nazionale per l’Informatica nella Pubblica Amministrazione, http://www.cnipa.gov.it/site/it-IT/
Towards a country-independent data format: the Akoma Ntoso experience.

Rete (Norms on the Network), aimed at providing a common data model for XML versions of national and regional norms. The NormeInRete schemas (NIR DTD and NIR XSD) and naming standards (NIR URN) have been published as national standards and are adopted by a number of local and national authorities, including the Italian Parliament and the Italian Court of Cassation. A brief description in English of the NormeInRete principles can be found in (Marchetti et al., 2002).

4.1. Simple document model

Data models created to handle complex document types (as legislation) need to deal with two apparently opposed requirements: on the one hand, they need to be sufficiently sophisticated to handle all possible occurrences and situations that may occur in the actual documents. On the other, they need to be speedily understood and used by the people who would need to apply these models.

These opposed requirements can be jointly satisfied not by simplifying the vocabularies of available structures and elements, which would reduce the available descriptive sophistication of the language, but rather by simplifying the structure variability and types (in XML lingo, the content models), thereby reducing the learning time and the software complexity without compromising a full and detailed descriptive power of the language. The idea therefore is to identify a number of basic, fundamental classes of structures (containers, hierarchies, blocks, etc.) that can be immediately understood and used appropriately, regardless of their actual names.

One of the main users considered when developing the AKOMA NTOSO model has been what we call the future toolmaker: this is a computer programmer that, fifteen years from now, will be asked to create new tools to manage AKOMA NTOSO documents or to activate new and currently unforeseeable computations with them. Differently from current toolmakers, the future toolmaker will not have access to complete documentation of the system, but only to sparse remaining documents, outdated by a fairly stratified situation where the basic ideas (on which the current toolmaker has worked) have evolved, modified, expanded and changed emphasis, often slowly and without documentation. The only reliable source of information available to the future toolmaker will be more than 15 years of actual legislation available in AKOMA NTOSO format.

Since the XML format has been created to be as complete and as self-

\footnote{NormeInRete, a joint project by the CNIPA and the Italian Ministry of Justice, \url{http://www.normeinrete.it/}}
explanatory as possible, we expect the future toolmaker, in principle, to deduce all undocumented facts about AKOMA NTOSO by simply examining a few relevant XML instances of the legislation and discovering there how it should work. In a sense, the future toolmaker is more a key user for our system than the current toolmakers, and the possibility for him/her to deduce fundamental properties of AKOMA NTOSO from the visual examination of XML documents is a guarantee of long-term existence and usefulness of the AKOMA NTOSO system itself.

Thus the AKOMA NTOSO model is by no means a one-time exercise. Given the quantity and diversity of Parliaments and Parliamentary documents under consideration, it is unavoidable that the AKOMA NTOSO model grows substantially over time. Besides being as much as possible self-explanatory, the AKOMA NTOSO model has been built to stand evolutions and changes regarding the number of actual functionalities provided: features such as the number of metadata, or the automatic generation of amended text, or the activation of special analysis tools on the text may require with time an evolution of the schema.

4.2. Presentation, Structure and Semantics

Modelling Parliamentary documents – particularly legislation – requires giving consideration to the interplay between presentation (how the information looks), structure (how the information is organized, e.g. in document parts) and semantics (what the information is/represents). The mainstream philosophy of descriptive markup languages (to whose family XML belongs) dictates that

− **semantic markup** is the richest and most complete, whereby fully semantically identified parts of the document (e.g., headings, names, references, etc.) imply both their structural role and the expected presentation,

− some **structural markup** may need to be added for those document parts (e.g., organization of parts, hierarchies, preambles, conclusions, etc.) that have no specific semantic role; on the other hand, the presentation related to this markup can also be implied by their structural role.

− Finally, **presentation markup** deals with purely typographical aspects of text: there might be some remaining text parts with no identifiable semantic or structural role, yet with a different presentation (e.g., a part of text that is in bold for no discernible
reason). These situations may occur because the current semantic/structural model is incomplete, or because the structure or semantics the part represents is a sporadic occurrence that was not deemed worth of being described, or because simply there was no real reason for putting a specific presentation but aesthetics. In these cases a full presentational model need to be described and used. Fortunately there are well known presentation models (e.g., XHTML + CSS) that can be directly applied to easily obtain full and sophisticated presentation effects.

There exists a continuum between explicit and hierarchical vocabularies that are specifics to a given domain, and flat, presentational vocabularies such as XHTML, as long as they allow descriptive information to be specified (e.g., the class attribute of XHTML elements) and allows unambiguous deductions as to the hierarchy they express.

This equivalence lets documents be created well across the semantic/presentational continuum, and still can be correctly converted into the final (fully descriptive) AKOMA NTOSO format automatically. For instance, it is be fairly easy to customize MS Word, Open Office, or any of a number of standard HTML editor, to produce texts with styles that can be automatically converted into the full AKOMA NTOSO format by deducing the actual names and hierarchy that the classes express.

4.3. Explicit and implicit structure identifiers

An issue whose support in AKOMA NTOSO can been seen as problematic is explicit naming. By design, AKOMA NTOSO needs to serve and deal with documents coming from several different African countries, where the names used for document parts may vary considerably from country to country, and where the same names can be used in different levels of the hierarchy.

A common practice, in SGML- and XML-derived languages, is to use tags that either reveal the name, or the structural role, of the parts being tagged. For instance, the hierarchical sections called “Part” and “Title”, would be rendered as <section> elements under an implicit naming approach, and as <part> and <title> under the explicit naming approach.

In AKOMA NTOSO we assume that the advantages of proper, explicit naming of objects are prevailing over the increased software complexity in managing multiple partially overlapping sets of element names. In fact, we believe that the added complexity is not very impractical to implement, and that the expected future advantages (especially for long-term understanding of structures) definitely overcomes the initial implementation disadvantages.
AKOMA NTOSO expects documents to use explicit names, yet at the same time allows different document types to use the same names in completely different contexts and order. This apparent difficulty is overcome by the adoption of two related families of schemas. Similarly to XHTML Transitional and Strict DTDs, in fact, AKOMA NTOSO defines one Generic Schema and as many related Detailed Schemas as needed.

The *Generic Schema* is designed as a very loose schema, containing the full elements’ vocabulary, but very few constraints on them; it allows all elements in all positions, and it is used as a baseline check for all situations in which no constraint can be imposed. Thus the GS validates all documents regardless of their provenance. Conversely, each individual *Detailed Schema* contains the same vocabulary, but imposes a larger number of constraints, compatible with the drafting rules of that specific document type. Detailed Schemas are subsets of the Generic Schema, and thus any document that is valid according to one of the DSs will also be perfectly valid according to the GS.

The only objection to this approach is that the Generic Schema also validates very absurd structures, and cannot be used to verify the structural correctness of documents. This is done on purpose: AKOMA NTOSO must accepts documents that were formally approved by the relevant Parliaments regardless of how absurd they happen to be. On the contrary, if a Parliament wishes to prevent the generation of absurd documents, it only needs to generate a Detailed Schema preventing inappropriate structures. This approach is more clearly detailed in the next section.

### 4.4. Verification of the validity of data

Validation is the act of checking the correctness of an XML document according to pre-defined structural rules expressed in one or more DTDs and XML Schemas. The validation verifies whether the XML document contains, in number and position, all the expected elements of the type this document is an instance of.

The problem with being too restrictive in the constraints of the AKOMA NTOSO schema is that the Parliaments may have approved, and may decide to approve in the future, documents that do not conform to these rules: although in most countries there are guidelines for the correct drafting of legislation, but this is just what they are: guidelines, that can be ignored and modified at will by a higher authority such as a Parliament.

This fact has a very important effect on the generation of XML versions of documents: everything that gets approved by Parliaments have to be
accepted by the system, and everything that has already been approved even more so. Therefore, failing XML validation (i.e., violating one or more of the constraints and restrictions expressed in the schemas) cannot have the effect of rejecting documents, but, at most, of pointing out issues and differences from the guidelines that the authority itself, if it wants and has time to spend on this, can consider for editing and modifications.

In reality there are two different actors in the complex issue of validating a piece of legislation: the Parliament, who is writing the actual content of the document, and the markers, who convert it into XML by identifying all interesting bits of the text and marking them up using the AKOMA NTOSO vocabulary.

In fact, we have designed the validation schema as a contract that only binds the marker, leaving the Parliament free to decide as it wishes. Thus compliance to rules such as “An identifier will always be added to each substructure of the act” or “The enactment date will be specified” can be safely required, as they bind the behaviour of the marker only, while structural rules (such as “Every subpart will have a heading”, or “A section will contain paragraphs which contain clauses”) cannot be imposed, as they would interfere with the authority and independence of the Parliaments.

Forcing markers to fully describe in XML all document parts, and yet leaving to the Parliament the maximum freedom in writing, may seem incompatible and hard-to-reach goals, but they can be and are reached in the AKOMA NTOSO framework. AKOMA NTOSO clearly separates data and metadata, thereby clearly distinguishing the contribution of the Parliament (data) and the contribution of the marker (metadata); AKOMA NTOSO provides a richly evocative vocabulary of structures and elements, so that the marker can correctly and precisely describe what is actually contained in the documents. AKOMA NTOSO imposes little or no constraints on data, letting the legislator write and organize the text matter as he wishes, but imposes a number of constraints on the metadata, forcing the marker of texts to provide all bits of information that are necessary to manage and organize the documents.

Yet it might be appropriate to also give guidance and help in following the drafting guidelines enacted in each country. This is the reason to provide both a General Schema and several Detailed Schemas: the GS is fully descriptive, only binding the marker and not the legislator, but allowing the marker to describe as precisely as possible the actual structure of the document as approved and generated by the Parliament. The DSs are more prescriptive, and are used to check whether the document actually conforms to the existing legal drafting guidelines in each individual country. Successful validation of documents will only
be required against the GS, as errors would signal incorrect markup from the marker, while the DSs can be used, at the discretion of the Parliament itself, to automatically check conformance of the proposed bill against the local drafting guidelines, and thus be able to modify it accordingly in case conformance is sought.

5. Design approaches in AKOMA NTOSO

The AKOMA NTOSO model deals with a rather complex situation of five document types and several African countries, by creating two classes of document types, the Generic Schema (GS) and several Detailed Schemas (DSs) that provide support for differences in document types. Interoperability across these schemas is granted by a generalized approach that maintains full descriptions of the element while unifying and limiting in scopes the structures.

This is obtained through the systematic use of patterns. Patterns are the abstraction and distillation of past experiences in designing and resolving design problems. They are general and widely applicable guidelines for approaching and justifying design issues that often occur in XML-based projects.

We distinguish between patterns in content models (a restriction of content models to the ones that are actually useful) and patterns in schema design (guidelines on how to make a schema more modular, flexible and understandable by users). Both types of patterns are well known and well established in the literature, although by different experts in different ways. For patterns in content models we rely on (Vitali et al., 2005), while the authoritative resource for patterns in schema design is (XML Patterns, 2006).

5.1. Support for multi-linguism

Due to the multi-language nature of the AKOMA NTOSO project, spanning across at least three European languages (English, French and Portuguese) and a number of African languages, all AKOMA NTOSO documents are required to use UTF-8 as their character encoding format.

Furthermore, although the schema has been designed to use English as the main language, it is expected that many users might have problems in using elements expressed in this language, and that relying on English alone would void the principle of descriptiveness of the documents.

For this reason, a (currently incomplete) list of equivalent names is provided in French and Portuguese and can be provided for any other
languages as well, that allow sophisticated tools to evaluate and use these documents as if they were written using the main English-based AKOMA NTOSO vocabulary.

5.2. DESIGN PATTERNS IN SCHEMA ORGANIZATION

Design patterns are distillation of common wisdom in organizing the parts and the constraints of a schema. AKOMA NTOSO refers systematically to patterns listed in (XML Patterns, 2006). Whenever there has been a design choice to be made that was not immediately obvious and naturally acceptable, a relevant pattern has been sought and properly used. In fact, a large number of patterns from (XML Patterns, 2006) have been used, but only a few of them need to be explicitly described and explained:

- **Universal Root**: One generic root element contains all elements describing the document types. The AKOMA NTOSO schema covers five different document types: acts, bills, parliamentary debate records, parliamentary order papers, and miscellaneous parliamentary documents. The naïve solution would be to create five different schemas, one for each document type. This would mean activating five different processes for editing, marking up, validating, converting and displaying these documents, all of which driven by information contained within the document itself, and unavailable in advance. By creating a single root element containing the actual document elements, on the other hand, we have a single schema describing correctly and completely all document types.

- **Consistent Element Set**: many elements share similar or identical internal structures (in XML lingo, they have the same content model). As mentioned in the previous section, all elements rely on five types of content models only, and all document elements share just two content models (a strictly hierarchical structure and a loosely hierarchical structure). The advantage of this approach is that the proliferation of element names does not detract from the overall clarity and simplicity of the overall schema.

- **Generic document & role attribute**: Besides named elements, the AKOMA NTOSO schema also provides for a generic element for each of the five main types of content model. These are elements called just like the corresponding pattern (thus they are called `<hierarchy>`, `<block>`, `<inline>`, `<marker>` and `<container>`), and are meant to be used for markup that fits the content model but for which no specific element was provided. A ‘name’ attribute
is provided for explicitly naming the element. A similar approach is taken in XHTML with the \texttt{<div>} and \texttt{<span> elements}, which are generic elements qualified by the ‘class’ attribute. It is also a characteristic of the AKOMA NTOSO schema that all named elements are equivalent, and can be substituted to the corresponding generic element with their name as the value of the name attribute. Thus, by definition, \texttt{<p>} is in AKOMA NTOSO the same as \texttt{<block name="p">}, \texttt{<part> is the same as \texttt{<hcontainer name="part">}}, \texttt{<act> is the same as \texttt{<container name="act">}}, etc.

- \textit{Reuse Document Types} (partial): many AKOMA NTOSO element types mean exactly the same as corresponding elements in XHTML. For this reason, rather than inventing new names that need to be learnt from scratch, correctly understood and actually used when marking documents, only to have them translated into the corresponding XHTML terms for display, we have decided to just use the XHTML name, saving time in explanation, documentation, learning and usage. Therefore, for instance, a table in AKOMA NTOSO uses the tags \texttt{<table>}, \texttt{<tr>}, \texttt{<td>} and \texttt{<th>} that were introduced in HTML, a paragraph is \texttt{<p>}, bold and italic are simply \texttt{<b>} and \texttt{<i>} respectively.

5.3. \textsc{Design patterns in content models}

The AKOMA NTOSO 1.0 Schema uses systematically five of the seven patterns described in (Vitali et al., 2005). This means that all content models and complex types used in the schema follow precisely the form of the relevant pattern, and all elements can be simply described and treated according to their pattern rather than individually.

These patterns are:

- The \textit{hierarchy}: a hierarchy is a set of arbitrarily deep nested sections with title and numbering. Each level of the nesting can contain either more nested sections or blocks. No text is allowed directly inside the hierarchy, but only within the appropriate block element (or, of course, titles and numbering).

- The \textit{blocks}: a block is a container of text or structures that is organized vertically on the display (i.e., has paragraph breaks) and can contain either substructures or text. Most blocks in AKOMA NTOSO are based on the HTML language.

- The \textit{inlines}: an inline element is an element placed within a mixed model element that identifies some text fragments as relevant for
some reason. There are both semantically relevant inlines and presentation oriented inlines. There is but one content model using inlines (and markers), which means that all mixed model elements (i.e., those that allow both text and elements) also allow the a repeatable selection of all inline elements.

- The markers: markers are content-less elements that are scattered here and there in the document and are meaningful for their names as well as their attributes. Markers are also known in literature as empty elements or milestones. There are two main families of markers in the AKOMA NTOSO schema: placeholders in the text content (e.g., note references) that can appear in any position that also has text, and metadata elements that only appear in some subsection of the <meta> section.

- The containers: containers are sequences of specific elements, some of which can be optional. Containers are all different from each other (as the actual list of contained elements vary), and so there is no single container content model, but rather a number of content models that share the record pattern.

The following is an example of a hierarchy of sections:

```xml
<clauses>
  <chapter id="chap2">
    <num>Chapter 2</num>
    <title>Traditional communities and...</title>
    <paragraph id="chap2-para2">
      <num>2</num>
      <title>Recognition of traditional...</title>
      <clause id="cla1">
        <num>1</num>
        <p>A community may be recognised as ...</p>
      </clause>
      ...
    </paragraph>
    ...
  </chapter>
  ...
</clauses>

5.4. Metadata elements

The meta section contains all the meta-information that needs or can be added to the actual content of the document. As a rule, all editorial content (i.e. content added by the editorial process out of Parliament
rooms) need to be placed in the meta section, except for markup and note references.

Meta elements are divided in four subsections:

− **Descriptors**: i.e., a set of metadata providing info about the document and its publication and edition details, including its official promulgation date, its official URI, and so on.

− **Lifecycle**: the lifecycle element provides information about the events that the document has undergone, and references to the documents that have caused these events.

− **Notes**: the notes element contains the text of the editorial notes that are produced to comment and expand the actual text of the document. Note references inside the text point to notes contained here.

− **Proprietary**: this subsection allows any additional metadata to be specified in any order and vocabulary (provided it uses a different namespace than AKOMA NTOSO). Proprietary metadata can be used within a specific document management system to provide additional information useful for internal search and document management that is not worth standardizing and imposing across all AKOMA NTOSO implementations.

The development of the meta section is not finished yet. For instance, support for Dublin Core metadata is currently imperfect (there are semantic equivalences between Dublin Core elements and AKOMA NTOSO elements, but they are not complete nor officially described as equivalent).

### 5.5. Identifiers

Identifiers are systematically used in AKOMA NTOSO. All AKOMA NTOSO elements allow an identifier. Many relevant elements and sections require it. Identifiers are the main way to identify fragments and parts of the document in an unambiguous form. They can be used in document references (e.g., links and amendment commands) as a precise pointer to the actual part of the document mentioned (as opposed to simply referring to a document as a whole). Also internal links need to use identifiers.

The AKOMA NTOSO schema also specifies a syntax for identifiers, composed by juxtaposing sub-identifiers of the path needed to access them. Legal documents provide explicit global numbering for sections and articles, and local numbering for hierarchical subparts of them. The following (Table I) is a table with some examples of identifiers:
Identifiers never change even if and when the elements get officially renumbered.

5.6. DOCUMENT URIs

A naming convention provides a simple mechanism for ensuring that Parliamentary Documents can be named and referred to in a consistent way. Being able to refer to documents in an unambiguous way that is both human and machine readable is very useful. In particular, a naming convention is critical in order to ensure that intra- and inter-Parliament hypertext linking can be achieved cost effectively.

All AKOMA NTOSO documents (and documents not converted into AKOMA NTOSO format, but referred to within a AKOMA NTOSO document) is identified by a unique name expressed as a specific URI. The actual syntax for AKOMA NTOSO URIs is being defined and is as yet unavailable.

6. Conclusions

The initial results of the AKOMA NTOSO framework are encouraging. Version 1.0 of AKOMA NTOSO is now complete, and contains schema and stylesheets for five different document types adapted to the needs of several different African Parliaments. Through the activity of UNDESA,
these schemas will soon evolve into local National Specifications for ten
African Parliaments that signed the Nairobi Declaration.

Given the importance of long term life span of the ideas behind
AKOMA NTOSO ideas, data and systems, it is vital that focused
activities for maintaining the project alive and active are taken. An
AKOMA NTOSO Management Board is being set up to ensure continu-
ing usefulness and adoption of the AKOMA NTOSO specifications, and
to engage in capacity building and users training at all levels (technical,
clerical, managerial) wherever the AKOMA NTOSO specifications have
been adopted, and to foster further adoption of the same.

References

Marchetti A., Megale F., Seta E., Vitali F., “Using XML as a means to access
legislative documents: Italian and foreign experiences”, ACM SIGAPP Applied

Vitali F., Di Iorio A., Gubellini D., Design patterns for document
substructures, Extreme Markup 2005 Conference, Montreal, 1-5 August
Vitali01/EML2005Vitali01.pdf

Vitali F., Zeni F., Working Towards Interoperability in African Parliamentary In-
formation Systems, IST-Africa 2006 Conference Proceedings, Paul Cunningham
and Miriam Cunningham (Eds). IIMC International Information Management

XML Patterns, http://www.xmlpatterns.com/