Chapter 3

XML Models for Legal Documents: Current Polish Practices and International Standards

Jerzy Duda
AGH University of Science and Technology

1. Introduction

The legislative processes of legal acts and their promulgation have undoubtedly entered a new area, which is based on the electronic form of creation, storage and information sharing. In the European Union member states are obliged to share public information by Directive 2003/98/EC on the re-use of public sector information (with amendments enacted in 2013). Currently, all countries provide access to national law in the form of HTML pages or PDF files, whereas a widely adopted and accepted format for information exchange on the Internet is presently based on the XML language. The main disadvantage of the traditional approach is the lack of semantic information about the real structure of a legal act and its content which makes a more advanced search in legal texts hardly possible, let alone the application of some knowledge engineering tools. Legal documents are stored locally in national databases, recorded in accordance with the individual, national data structure, and this further complicates the reuse of such documents, especially in a multinational environment. The solution to the above problems is the application of XML language in national legislative processes and the development of a common transnational standard describing the structure of XML documents from national legal systems. It is possible, since in many countries including Austria, Denmark, Italy, the Netherlands, Spain and Poland, legal documents are already prepared in XML format, while similarly there exist some transnational XML standards like Metalex or Akoma Ntoso that can be

used in order to describe the structure of national legal acts. The goal of the paper is to discuss the problem of the development of a single national XML standard for legal purposes and its compatibility with transnational standards on the basis of current practices in Poland.

In the following sections basic concepts related to the application of XML language in the legal domain are described, then the most noteworthy examples of XML Schemas for national and transnational legal systems are presented and discussed. Finally the development of an XML legal standard in Poland is analyzed, with special respect to the compatibility of the proposed solutions with transnational standards.

2. XML language and related tools

XML language is currently regarded as the most universal language for data description and sharing in an electronic format. Primarily designed mostly for data exchange between different Internet applications,[1] it has also slowly become a common language for writing various types of documents, including documents created in popular office suites (Open Office, MS Office). The widespread acceptance of the XML standard, almost immediately after its release in 1998,[2] contributed to the postulate of its application also for the documents related to public administration. It is worth noting that XML is not an entirely new concept. The first language that allowed the marking up of a document with tags describing its structure (like headers, paragraphs, tables and so on) was the Generalized Markup Language (GML), developed in the 1960s for IBM. Its descendant – Standard Generalized Markup Language (SGML), originally aimed at sharing documents in large governmental, legal and industrial projects in a human readable form,[3] and approved in 1986 as an ISO norm,[4] became the first publicly available and widely used markup computer language. Both HTML and XML standards are in fact the subsets of SGML standard, much more restricted than its predecessor and targeted for a specific application.

---

[1] R. Jaideep, R. Anupama, *XML: Data's Universal Language*, IT Professional, May–June 2000, Vol. 2, No. 3, p. 32–36.

[2] XML 1.0 Recommendation was issued by World Wide Web Consortium (W3C) in February 1998, however, the Working Draft has been published earlier, in November 1996.

[3] V. Kashyap, Ch. Bussler, M. Moran, *The Semantic Web: Semantics for Data and Services on the Web*, Berlin, Heidelberg 2008, p. 293.

[4] ISO 8879:1986 Information processing – Text and office systems – Standard Generalized Markup Language.

2.1. Comparison between XML and HTML

The fundamental language for information publishing in the Internet network – HTML – was proposed by Berners Lee in 1991 as a part of an Internet-based hypertext system and originally consisted of 18 tags that could be interpreted by a web browser. In HTML (contrary to XML) tags cannot be extended by a user, so the real structure of the documents cannot be described properly. Fig. 1 depicts an excerpt from the HTML document containing the Constitution of the Republic of Poland published on the website of Polish Parliament.

```
<HTML>
 <HEAD>
  <META HTTP-EQUIV="Content-type" CONTENT="text/html; charset=iso-
8859-2">
   <TITLE>The Constitution of the Republic of Poland</TITLE>
   <link href="../../../style1-2.css" REL="stylesheet"
   TYPE="text/css">
 </HEAD>
 <body topmargin="2">
   <br>
     <BLOCKQUOTE>
     <P ALIGN="CENTER"><FONT COLOR=#A00000><B>THE CONSTITUTION OF THE
     REPUBLIC OF POLAND<BR> OF 2nd APRIL, 1997</B></FONT></P>
     <P ALIGN="CENTER"><B>As published in <I>Dziennik Ustaw </I>No. 78,
     item 483</B></P> [...]
     <B><P ALIGN="CENTER">Chapter I<BR>
     <FONT COLOR=#101084>THE REPUBLIC</FONT></P></B>
     <I><B><P ALIGN="CENTER">Article 1</P></B></I>
     <P>The Republic of Poland shall be the common good of all its
     citizens.</P>
     <B><I><P ALIGN="CENTER">Article 2</P></B></I>
     <P>The Republic of Poland shall be a democratic state ruled by law
     and implementing the principles of social justice. </P>
     <B><I><P ALIGN="CENTER">Article 3</P></B></I>
     <P>The Republic of Poland shall be a unitary State.</P> [...]
   </BLOCKQUOTE>
  </BODY>
</HTML>
```

Fig. 1. First three articles of the Constitution of the Republic of Poland written in HTML Source[5]

The HTML document presented above contains tags describing the basic structure of the document: quotation <BLOCKQUOTE>, paragraph <P>, ordered list <OL>, and list items <LI>, along with the text formatting tags, informing a browser how to display the contents: bold <B>, italic <I> and

---

font with its color attribute <FONT COLOR>. However, neither the original structure of the legal document is preserved, nor the full description of the legal act is provided.

The main advantage of using XML instead of HTML is the possibility of using any set of tags that can have human readable names and precisely describe the original structure of a document. Fig. 2 presents the same fragment of the Constitution of the Republic of Poland as before, but this time written in XML language.

```xml
<?xml version="1.0" encoding="utf-8"?>
<act>
    <description status="in force" published="Dziennik Ustaw No. 78,
Item
     483" issue-date="1997-04-02" entry-date="1997-10-17">
     <title> The Constitution of the Republic of Poland</title>
    </description>
    <chapter no="I">
     <title>THE REPUBLIC</title>
      <article no="1">
        The Republic of Poland shall be the common good of all its
      citizens.
      </article>
       <article no="2">
        The Republic of Poland shall be a democratic state ruled by law
      and implementing the principles of social justice.
      </article>
       <article no="3">
        The Republic of Poland shall be a unitary State.
      </article>
        [...]
</act>
```

Fig. 2. First three articles of the Constitution of the Republic of Poland written in XML
Source: own elaboration

Storing a document in an XML format has even more advantages. First of all, there are many tools that can be used in order to process its content for the purposes of knowledge extraction or transformation to any other form of presentation required by the final user. Moreover, XML documents are natively ready for internationalization as they use Unicode standard for character encoding.

2.2. Definition of XML document

The structure of an XML document is usually strictly defined. An XML document, that contains a related definition of its structure (usually in a separate file), can be checked for its "validity".[6] This means that all tags and their attributes have been properly used, i.e. names of the tags are correct, they occur in the right order and are properly nested. Tag attributes are also checked for their names and values. Depending on the document definition standard, other restrictions can be applied for both tags and attributes.

Originally, a document type definition (DTD) standard, based on SGML, was widely used in order to define the structure of XML documents. Fig. 3 shows DTD definition for the fragment of the constitution presented in Fig. 2 (more detailed sections of a legal act, such as paragraphs, points, letters, were intentionally omitted).

```
<!ELEMENT act (description, chapter+)>
<!ELEMENT description (title)>
<!ELEMENT title (#PCDATA)>
<!ELEMENT chapter (title, article+)>
<!ELEMENT article (#PCDATA)>
<!ATTLIST description status CDATA #REQUIRED>
<!ATTLIST description published CDATA #REQUIRED>
<!ATTLIST description issue-date CDATA #REQUIRED>
<!ATTLIST description entry-date CDATA #REQUIRED>
<!ATTLIST chapter no CDATA #REQUIRED>
<!ATTLIST article no CDATA #REQUIRED>
```

Fig. 3. Exemplary DTD schema for a legal act
Source: own elaboration

DTD structure, despite the fact that it uses a specific format of tags, can be relatively easily read by humans. For example, from the DTD document presented in Fig. 3, the hierarchy of sections in a legal act can be read: An act has to contain a description and at least one chapter. A chapter must possess a title and at least one article. An article can contain either data or paragraphs, and finally, paragraphs (if present) contain data. All of the following attributes are required: status, date of issue and date of entry into force for the act's description, and the numbers for the act's subsections.

An XML structure definition in DTD is simple but primarily lacks definition of data types (e.g. a date can be written as any combination of alphanu-

---

[6] XML documents are checked in two phases. First the document is checked to see whether it is well-formed, i.e. if tags are closed, if they are properly nested, and if the document begins with a single root tag. Next, a well-formed document can be further checked as to whether it is valid against a given structure definition.

merical characters that have no sense in terms of description of some point in time), does not allow for namespaces (e.g. an act as a legal document means something different than act as an action – if they exist in a single XML document without namespaces their meaning is ambiguous). The XML Schema (XSD) proposed by W3C in 2001 was meant to be devoid of these shortcomings, and first of all, it was based on XML language. Fig. 4 shows a definition of a legal act in XML Schema, analogous to the DTD, presented in Fig. 3.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema">
    <xs:element name="act">
       <xs:complexType>
          <xs:sequence>
             <xs:element ref="description"/>
             <xs:element ref="chapter" maxOccurs="unbounded"/>
          </xs:sequence>
       </xs:complexType>
    </xs:element>
    <xs:element name="description">
       <xs:complexType>
          <xs:sequence>
             <xs:element ref="title"/>
          </xs:sequence>
          <xs:attribute name="status" type="xs:string"
use="required"/>
          <xs:attribute name="published" type="xs:string"
use="required"/>
          <xs:attribute name="issue-date" type="xs:date"
use="required"/>
          <xs:attribute name="entry-date" type="xs:date"
use="required"/>
       </xs:complexType>
    </xs:element>
    <xs:element name="chapter">
       <xs:complexType>
          <xs:sequence>
             <xs:element ref="title"/>
             <xs:element ref="article" maxOccurs="unbounded"/>
          </xs:sequence>
          <xs:attribute name="no" type="xs:string" use="required"/>
       </xs:complexType>
    </xs:element>
    <xs:element name="article">
       <xs:complexType>
          <xs:simpleContent>
             <xs:extension base="xs:string">
                <xs:attribute name="no" type="xs:byte"
use="required"/>
             </xs:extension>
          </xs:simpleContent>
       </xs:complexType>
    </xs:element>
</xs:schema>
```

Fig. 4. Exemplary XML Schema for a legal act
Source: own elaboration

While allowing for a more precise definition of XML document structure, XML Schema is simultaneously much more complex than DTD. This is why DTD is still used in many cases for the purpose of structure definition of XML files, including the definition of legal documents. In the next section, examples of both DTD and XML Schema used for definition of a legal documents structure will be presented.

2.3. Presentation layer for XML documents

Contrary to the HTML document shown in Fig. 1, the constitution written in the XML document, shown in Fig. 2, contains no tags related to the formatting of the document.[7] It means that an XML document should be accompanied by a file that describes how the XML content must be formatted for display on screen or for printing. This is usually done using XSLT (XSL Transformation) language, which belongs to the wider family of XML tools called XSL (Extensible Stylesheet Language). By executing a script written in an XSLT sheet, the XML document can be transformed into another file format, usually HTML, or indirectly, to PDF file. It is worth noting that an XML document may contain a direct reference to the XLST sheet, and a web browser (or other tool) can use it in order to display the content of XML file, instead of simple displaying XML code, which is the default behaviour. Fig. 5 illustrates a simple XSLT script that produces HTML document on the basis of XML file. The individual sections are formatted in a similar way to the HTML document, presented earlier in Fig. 1.

The XSL family also has yet another language for formatting the content of an XML document which is called XSL-FO (XSL – Formatting Objects). Contrary to the XSLT instructions, XSL-FO tags are placed directly into XML document. This is usually done by executing a specially prepared XSLT sheet after which the XML file must be processed by a special processor, e.g. Apache FOP (Formatting Objects Processor). The final output of this process can be a PDF file, MS RTF, MS Word, or another document format. The process of converting an XML document into PDF format is illustrated in Fig. 6.

---

[7] Actually, an HTML (or XHTML) document prepared according to the current rules of web pages design should also contain no tags related to formatting, as the presentation layer should be moved entirely to the CSS style sheet written in a separate file.

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
   <xsl:template match="act">
      <html>
      <head> <title/> </head>
      <body>
         <p style="text-align:center;font-weight:bold;color:#A00000">
            <xsl:value-of select="upper-
case(description/title)"/></p>
         <p style="text-align:center;font-weight:bold">
            <xsl:text> As published in </xsl:text>
            <xsl:value-of select="description/@published"/></p>
         <xsl:for-each select="chapter">
            <p style="text-align:center;font-weight:bold">
               <xsl:text>Chapter </xsl:text>
               <xsl:value-of select="@no"/><br/>
               <span style="color:#101084">
               <xsl:value-of select="upper-case(title)"/></span></p>
               <xsl:for-each select="article">
                  <p style="text-align:center;font-weight:bold;
                     font-style:italic">
                     <xsl:text>Article </xsl:text>
                     <xsl:value-of select="@no"/></p>
                  <p><xsl:value-of select="."/></p>
               </xsl:for-each>
         </xsl:for-each>
      </body>
      </html>
   </xsl:template>
</xsl:stylesheet>
```
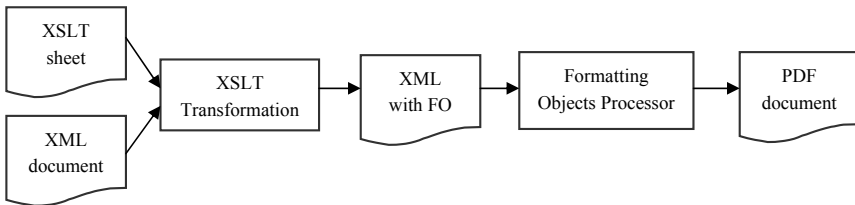
Fig. 5. Exemplary XSLT sheet transforming part of legal act into HTML document
Source: own elaboration

Fig. 6. Conversion process of XML document into PDF file
Source: own elaboration

In a typical scenario, an XML file with embedded formatting objects is intended to be a temporary file, as a result of the transformation of XML document, and this file is usually deleted after the conversion process is finished. The syntax of XML FO vocabulary is complex and this causes XML

documents with FO tags to be barely readable by humans. A simple example of such a file that contains the tags responsible for formatting only the page layout and the title section of the legal act is shown in Fig. 7.

```
<?xml version="1.0" encoding="UTF-8"?>
<fo:root xmlns:fo="http://www.w3.org/1999/XSL/Format">
   <fo:layout-master-set>
      <fo:simple-page-master master-name="page" margin-left="2cm"
       margin-right="2cm" page-height="29.70cm" page-width="21cm"
       margin-top="3cm" margin-bottom="3cm">
         <fo:region-body margin-top="1cm" column-count="1"/>
      </fo:simple-page-master>
   </fo:layout-master-set>
   <fo:page-sequence master-reference="page">
      <fo:flow flow-name="xsl-region-body">
         <fo:block font-weight="bold" color="#A00000" text-
         align="center" space-after="5mm">
            THE CONSTITUTION OF THE REPUBLIC OF POLAND
         </fo:block>
         <fo:block font-weight="bold" text-align="center">
            As published in <fo:inline font-style="italic">Dziennik
            Ustaw</fo:inline> No. 78, Item 483
         </fo:block>
      </fo:flow>
   </fo:page-sequence>
</fo:root>
```

Fig. 7. An example of XML file with embedded formatting objects
Source: own elaboration

The main difference between XML and HTML languages is that the former was developed in order to describe its contents while the latter was mainly intended for hyperlinking and presentation. The HTML that is still used in many legal information systems contains no information (at least not in a direct form) related to the structure of the document, or any other additional semantic information (unless it is somehow hidden).[8] An XML document, in turn, usually lacks any information about the presentation layer, thus it should be accompanied with some XSLT file, able to produce the desired output. If the XML documents are intended to be shared among various users (and legal acts definitely are), they should also follow the strict rules of syntax that are defined in DTD or XML Schema file. Otherwise, the original message written by a legislator may be decoded in an incomplete or even wrong way by recipients.

---

[8] HTML or XHTML file may use class attributes for an indirect structure description. However, class attribute is intended to be used as a selector for CSS, not for semantic purposes.

3. XML language for legal drafting and publishing

Historically, the first legal information systems were created on the basis of popular formats like HTML, RTF, MS Word .doc, ASCII files, and later PDF format, stored directly in a system's database or were generated on the fly from data stored in internal format in such databases (e.g. SQL tables, flat files). For example, in the Legal Information System of the Republic of Austria (RIS[9]) created in 1983, legal acts can be retrieved in HTML, RTF and PDF format. Legal drafts are prepared in MS Word format and stored in a dedicated legislation system called E-Recht. A special plugin for the MS Word editor enables legislators to authenticate to the E-Recht system and to convert Word documents into the XML files that are finally stored in the system. In the future, the XML format will also be used as a retrieval format.[10]

In Europe, the XML standard is currently reported to be used in legal information systems in the Netherlands (SDU BWB), Italy (Norme-In-Rete), Dania (LexDania), Switzerland (CHLexML),[11] and in recent years also in Spain,[12] Estonia, Germany and Ireland.[13] Also European Union Publication Office uses the FORMEX system which, since 2004, has been based on XML language. In the following subsection, the most noteworthy standards will be described.

3.1. National XML standards for legal purposes

The Basiswettenbestand (BWB) legal system contains almost all of the regulations of Dutch law and currently uses the XML format to store legal texts in its database. XML documents must comply with the BWB DTD schema.[14] This schema defines elements able to describe the structure of four main types of

---

[9] Das Rechtsinformationssystem des Bundes.

[10] The e-law project in Austria Electronic support of law-making, Parliamentary Administration Engeljehringer – Schefbeck, Vienna, February 15, 2006, http://www.parlament.gv.at/ZUSD/PDF/2006-04-18_Publikation-Englisch.pdf (accessed 21.06.2013).

[11] An extensive review of XML schemas used in these countries can be found in: C. Lupo et al., *General XML format(s) for legal sources – Estrella European Project IST-2004-027655*. Deliverable 3.1, Faculty of Law, University of Amsterdam, Amsterdam 2007.

[12] D. Blanco, V. Javier, M. González, M. Mercedes, *Spain on Going Legislative XML Projects*, in: C. Biagioli, E. Francesconi, G. Sartor (eds*.), Proceedings of The V Legislative XML Workshop*, European Press Academic Publishing, February 13, 2007, p. 23–38.

[13] LEOS: Legislative Editing Open Software. Final Results, The European Commission ISA, May 2012, https://joinup.ec.europa.eu/sites/default/files/ISA_LEOS_Final_Results_Final_Version.pdf (accessed 21.06.2013).

[14] Current version of the BWB DTD is available at: http://koop.overheid.nl/producten/basis-wetten-bestand/documentatie (accessed 21.06.2013).

Dutch legal documents: law (*wet*) and decision (*besluit*), regulation (*regeling*), circulaire and international agreement (treaty – *verdrag*). The structure of the common element containing either law or decision is shown in Fig. 8.
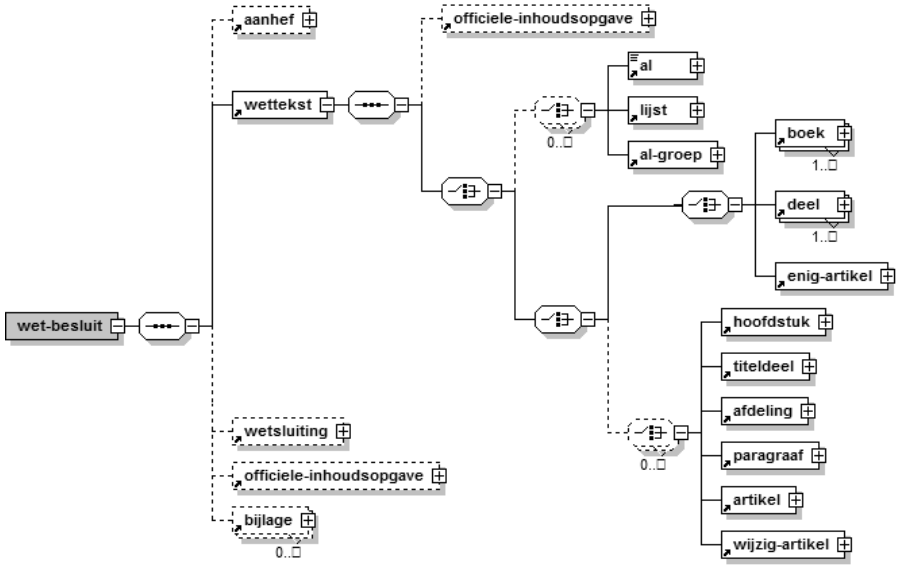

Fig. 8. Schema of *wet-besluit* element from Dutch BWB DTD schema
Source: own elaboration[15]

The main structure of a Dutch legislation document includes: preamble (*aanhef*), the text of law (*wettekst*), closure (*wetsluting*), table of contents (*officiele-inhoudsopgave*) and optional appendices (*bijlage*). The text of law may be organized within books (*boek*), parts (*deel*), chapters (*hoofdstuk*), titles (*titeldeel*), divisions (*afdeling*) and sections (*paragraaf*).[16] Each section may not strictly follow the whole hierarchy and contain articles (*artikel*). Articles can also be written directly without any section. The structure of a simple article in BWB XML consists of paragraphs (alinea – *al*) and lists (*lijst*). The paragraphs next to the text (#PCDATA) may also contain some formatting elements like emphasis (in attributes the exact formatting like bold, italic or caps maybe set), superscript, subscript, underline, and overline.

---

[15] Visualization prepared in AltovaXML Spy (Home) on the basis of BWB XML DTD schema.

[16] E. de Maat, R. Winkels, T. van Engers, *Making Sense of Legal Texts*, in: *Formal Linguistics and Law 212*, G. Grewendorf, M. Rathert (eds.), Berlin 2009, p. 225–255.

BWB XML also contains tags related to the legislation process and modi-
fication of the legal document or its parts. Another group of tags describe the
publishing process. In order to manage such a vast number of tags, elements
in BWB DTD were grouped into 19 different modules.

A more complex approach to XML schema was applied in the Italian No-
me-In-Rete system. The schemas are organized into separate files. There are
two DTD schemas used for the definition of a legal document structure –
a flexible version (*nirflessibile.dtd*), that is dedicated to the legal acts not fol-
lowing the drafting rules in a circular published on the 20th April 2001,[17] and
a strict version (*nircompleto.dtd*), dedicated to a new law. The strict version is
more restrictive than the flexible version, but a downward compatible. XML
document that is valid according to the flexible DTD schema, is also valid ac-
cording to the strict DTD schema. Fig. 9 shows the element *NormaNazionale*
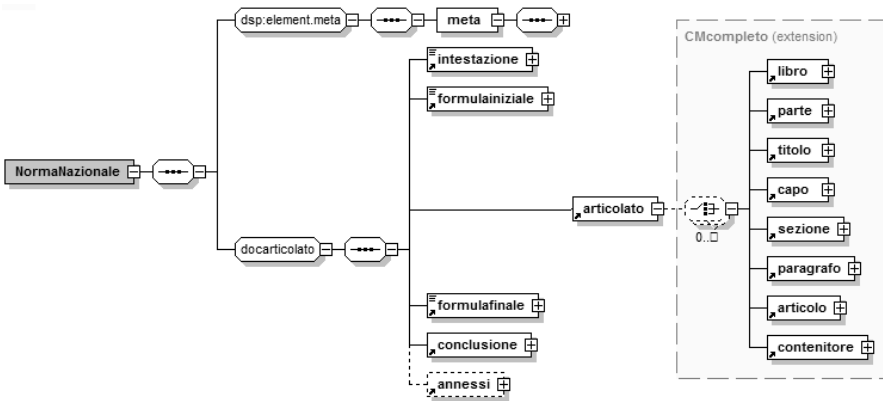that describes a national provision in NIR.

Fig. 9. Schema of *NormaNazionale* element from Italian NIR DTD schema
Source[18]

A national provision (constitution, legislative decree, statutory instru-
ment, regional decree, etc.) is structured with some meta information, intro-
duction (*intestazione*), initial formula (*formulainiziale*), the main articulated
content (*articolato*), ending formula (*formulafinale*), conclusion (*conclusione*)
and optional appendices (*annessi*). In the strict DTD version, the element
articolato has to contain sections, that are further organized in a strict hierar-

---

[17] CIRCOLARE 20 aprile 2001, n. 10888, Regole e raccomandazioni per la formulazione
tecnica dei testi legislative (Gazzetta Ufficiale n. 97 del 27.04.2001).

[18] Visualization prepared in AltovaXML Spy (Home) on the basis of CIRCOLARE 20
aprile 2001, n. 10888.

chical order. In the flexible DTD element *articolato* may also be divided into sections, but the strict use of sections (and their order) is not required.

The *nircompleto.dtd* and *nirflessibile.dtd* DTD schemas (in the second version of NIR renamed to *nirstrict.dtd* and *nirfexible.dtd*, respectively) depend on the set of other DTD schemas. All the elements related to the structure of legal act (including *NormaNazionale*) are defined in *norme.dtd*. Among all other DTD schemas, the most important are: *globali.dtd,* that contains the definitions of attributes and elements used throughout other schemas, *testo.dtd*, that contains the formatting elements compatible with HTML, and other tags related to the visualization of a legal document, and finally *meta.dtd,* that contains elements used in the metadata section of the document. Metadata are divided into over thirty groups that include elements related to, for example, the drafting process (*lavoripreparatori*), editorial notes (*redazionale*), publication (*pubblicazione*) and, in NIR2, also related to the life cycle management of the legal document (*ciclodivita*): events (*eventi*) and relations (*relazioni*).[19]

The schemas for Norme-in-Rete are also currently available in the format of XML Schema that are fully compatible with DTD.[20]

3.2. Transnational XML legal standards

In 2000, on the crest of a wave of tremendous interest in the development of the e-economy and creation of standards relating thereto, the OASIS Group, one of the major organizations developing XML standards, established a group responsible for the creation of a common XML standard for various legal purposes called LegalXML. In July 2000 there were over 545 Legal XML members from private companies (50%), government (25%), academic (12,5%), and non-profit organizations (12,5%), coming primarily from the United States.[21] In Europe members were organized within the LEXML initiative, which was established at the legal session of XML Europe conference held in Berlin 2001 (supported mainly by academics from Austria, Germany, Netherlands, Spain, Sweden, and later also Brazil).[22] Many different workgroups[23] were created within OASIS, each responsible for the development

---

[19] Compare: C. Lupo et al., *General XML format(s) for legal sources...*, p. 22–33.
[20] Linee guida per la marcatura dei documenti normativi secondo gli standard Normeinrete, Agenzia per l'Italia digitale, European Press Academic Publishing, Florence 2012; http://www.digitpa.gov.it/sites/default/files/LineeGuida_marcatura%20documenti%20normativi.pdf (accessed 21.06.2013).
[21] "Todd" V. Winchel, III, *Legal XML and Standards for the Legal Industry*, SMU Law Review, 2000, Vol. 1395, p. 1397–1404.
[22] http://www.juridicum.su.se/lexml/lexml/index.htm (accessed 21.06.2013).
[23] Called formally Technical Committees (TC).

of an XML standard for a different application in the legal domain: Court Filing, eContracts, eNotary, IntJustice (Integrated Justice – for exchanging data in civil and criminal cases), Lawful Intercept, Legislative, ODR (On-line Dispute Resolution) and Transcripts.[24] From all of these projects the final XML standards were published only for electronic court filling and electronic contracts. All other projects were postponed.

In 2002 a project called E-POWER was launched, initiated by the Leibniz Center for Law of the University of Amsterdam, in cooperation with Application Engineers and the Dutch Tax and Customs Administration (DTCA), and sponsored by the European Union.[25] One of the outcomes of this project was a MetaLex standard, aimed at being a language and jurisdiction-neutral standard for legislative purposes. The most recent version of Metalex was published in 2010 by the European Committee for Standardization (CEN) as CEN Workshop Agreement no 1571.[26]

The elements of Metalex language are used primarily for the description of a legal documents structure. The current specification of the language includes five types (models) of document content. Text content, similar to HTML, can be written in block elements or inline elements, wherein inline elements can be surrounded by block elements, but not the opposite. Block elements can contain empty elements marking a position in text (*milestone*). Text elements can be grouped into hierarchic container elements (*hcontainer*), that also include a title (*htitle*). The document begins with the root element. The root element and container elements may also act as a container for metadata (*mcontainer*). The structure of the root element in Metalex is shown in Fig. 10.

All elements representing different content type are instances of accompanying abstract elements that can be further extended according to the national legislative system. If the legal act contains, for example, articles in the Metalex language, they can be defined as an instance of a hierarchical container. An example of such a definition that preserves the restrictions defined originally for *hcontainer*, written in XML Schema, is provided in Fig. 11.

As a part of the Estrella project, an XSLT transformation tool for conversion from an Italian NIR DTD to MetaLex was prepared (NIR2Metalex). All 273 NIR elements were converted into Metalex elements, while also metadata were transformed into RDFa attributes. Some special containers, which were specific for NIR, had to be exclusively defined, e.g. general metadata (*gmeta*) for "disposizioni" type of metadata. These metadata can contain text, content elements and other metadata.

---

[24] http://www.legalxml.org/ (accessed 21.06.2013).

[25] http://www.metalex.nl/ (accessed 21.06.2013).

[26] Metalex (Open XML Interchange Format for Legal and Legislative Resources), ftp://ftp.cen.eu/CEN/Sectors/List/ICT/CWAs/CWA15710-2010-Metalex2.pdf (accessed 21.06.2013).

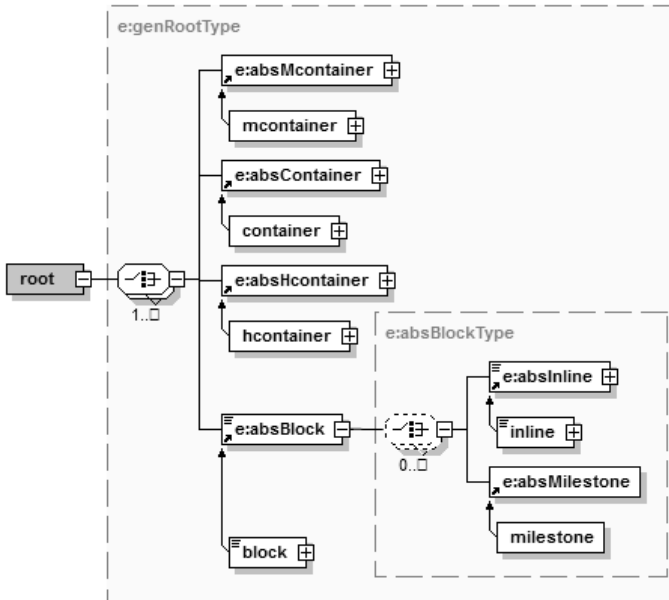**Fig. 10.** The structure of root element in Metalex
Source[27]

```
<xsd:complexType name="articleType">
    <xsd:complexContent>
        <xsd:restriction base="e:hcontainerType">
            <xsd:sequence>
                <xsd:element ref="e:absHtitle" maxOccurs="unbounded"/>
                <xsd:choice>
                    <xsd:element ref="e:absHcontainer"
maxOccurs="unbounded"/>
                    <xsd:element ref="e:absContainer"
maxOccurs="unbounded"/>
                </xsd:choice>
            </xsd:sequence>
            <xsd:attributeGroup ref="e:globnumb"/>
        </xsd:restriction>
    </xsd:complexContent>
</xsd:complexType>
```

Fig. 11. XML Schema fragment defining an article in MetaLex (old version)
Source[28]

---

[27] Visualization prepared in Altova XML Spy on the basis of Metalex schema available at: http://justinian.leibnizcenter.org/MetaLex/e.xsd (accessed 21.06.2013).

[28] E. de Maat et al., *Translators. (Extended) tools for translating existing other formats into new format*, Estrella European Project IST-2004-027655. Deliverable 3.3, Faculty of Law, University of Amsterdam, Amsterdam 2007, p. 9–10.

For the Dutch law, a system of automatic conversion was developed by Hoekstra.[29] The structure of BWB documents had to be defined using Metalex content type elements, and mapping schema for the elements existing in BWB DTD to the new structure had to be provided. This step was done in a semiautomatic way (some elements in BWB structure had to be defined manually). Finally, the legal documents written in BWB XML could be automatically transformed into the Metalex format, and stored on the server.

In 2004–2005 another transnational legislative language called Akoma Ntoso[30] was developed as a part of the "Strengthening Parliaments' Information Systems in Africa" project, maintained by UN Department for Economics and Social Affairs. Akoma Ntoso was aimed at the development of common standards for parliamentary, legislative and judiciary documents[31] for the Pan-African interoperability of Parliaments. The schema contains one universal root element (*akomaNtoso*), that describes all types of documents: act, bill, judgment, debate, amendment, statement, and other documents and their aggregations or summaries: official gazette, document collection, amendment list, debate report. Some of them (e.g. acts and bills) can have hierarchical structure, some of them (e.g. document, statement) can have an open structure (elements can be included inside other elements in almost any order), while others have the structure related to a particular document type (e.g. amendment or judgment). Fig. 12 shows the structure of act element in Akoma Ntoso, used for legislative purposes. It has a typical, predefined hierarchical structure (similar to the BWB XML or NIR XML structure), that includes: meta, cover page, preface, preamble, the main body of the act, conclusions and attachments.

The body part of the document can use six different patterns for describing its content: *markers* (context-less elements similar to *milestone* in MetaLex), patterns analogous to MetaLex: *inlines* (both semantically relevant inlines and presentation oriented, *blocks* (Akoma Ntoso blocks are based mainly on HTML tags), *containers* (organizing blocks and other containers), and finally *popups* (elements that are out of current hierarchical context, used in mixed model), and *hierarchy* (only one hierarchy with 21 predefined hierarchical elements is used in Akoma Ntoso version 3.0).

---

[29] R. Hoekstra, *The MetaLex Document Server – Legal Documents as Versioned Linked Data*, Proceedings of the 10th International Semantic Web Conference ISWC 2011, Berlin––Heidelberg 2011, p. 128–143.

[30] Akoma Ntoso is an abbreviation for Architecture for Knowledge-Oriented Management of African Normative Texts using Open Standards and Ontologies, in Akan language of West Africa it means "linked hearts".

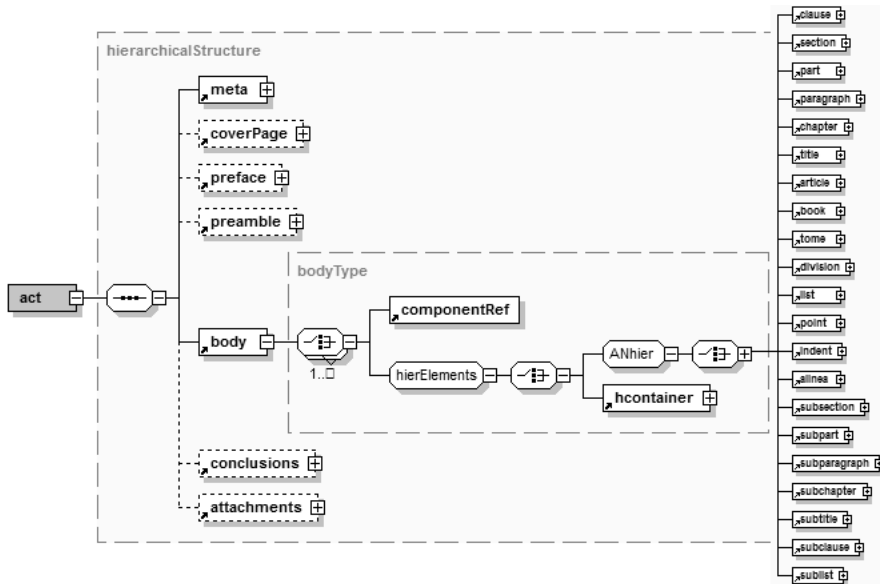[31] http://www.akomantoso.org/akoma-ntoso-in-detail/scope (accessed 21.06.2013).

Fig. 12. The structure of act element in Akoma Ntoso
Source[32]

The Akoma Ntoso standard was successfully applied by African Parliaments, but also customized for the legislative purposes of Brazilian Senate (LexML) and European Parliament (XML4EP editor).[33] It also became the basis for OASIS LegalDocML, a new initiative started in 2012, one of whose main goals is "to create a common legal document standard for the interchange of parliamentary, legislative and judicial documents between institutions anywhere in the world".[34]

4. Electronic form of legal acts in Poland

In Poland, the use of ICT in public sector is regulated by the Act of 17 Feb. 2005 on informatization of entities performing public tasks (Dz.U. 2005, no. 64, item 565, as amended). This act adds a new article to the Act of 20 July 2000 on the publication of legislative acts and of certain other acts

---

[32] Visualization prepared in Altova XML Spy on the basis of Akoma Ntoso scheme available at: http://www.akomantoso.org (accessed 21.06.2013).

[33] M. Palmirani, F. Vitali, *Akoma-Ntoso for Legal Documents, Legislative XML for the Semantic Web Law*, Governance and Technology Series, 2011, Vol. 4, p. 75–100.

[34] A new standard should also be compatible with CEN/MetaLex, https://www.oasis-open.org/committees/legaldocml/charter.php (accessed 21.06.2013).

(Dz.U. 2000, no. 62, item 718), stating that "official journals, collections of local law […] and other normative legal acts shall be made available free of charge by means of electronic communication" and that "minister in charge of information technology shall determine, by regulation, the technical requirements to be met by means of electronic communication and data carriers used to provide official journals, collections of local law […]". The Act of February 17, 2005 also introduces the concept of: 1) the minimum requirements for ICT systems used to perform public tasks, 2) the concept of the minimum requirements for public registers and exchange of information in electronic form, and establishes the rules for determining these requirements. Moreover, the act states that these requirements should ensure the openness of information standards. In Appendix 2 to the subsequent Regulation of The Council of Ministers of October 11, 2005 on the minimum requirements for IT systems (Dz.U. 2005, no. 212, item 1766), covering the data formats to provide access to information resources made available through IT systems used to perform public tasks, the following formats determining the structure of the information in an electronic document are listed: XML, XSD (XML Schema), GML (Geography Markup Language used in GIS), and for the transformation of electronic documents: either XSL or XSLT. The newest regulation[35] additionally allows for RELAX NG standard[36] of XML structure description, but first of all, requires technological neutrality of the solutions.

4.1. XML legal standardization in Poland

The Regulation on the technical requirements of electronic documents containing normative acts and other acts, electronic form of official journals and electronic means of communication and information media was finally issued by the Minister of Internal Affairs and Administration on April 25, 2008 (Dz.U. 2008, no. 75, item 451, as amended). According to this regulation, "Electronic documents containing acts being subject to announcement shall be prepared in XML format […]" and "Acts being subject to announcement in the official journals, collections of local law constituted by district or the municipality shall be published in XML and PDF formats based on XML document". In Annex 5, the detailed XML Schema was provided.[37] Later this schema was updated in 2009–2010, along with the editor of legal acts in XML (called EDAP), developed at the request of the Ministry of Internal Affairs

---

[35] Regulation of Council of Ministers of May 16, 2012 on the National Interoperability Framework, the minimum requirements for public registers and exchange of information in electronic form, and the minimum requirements for IT systems (Dz.U. 2012, no. 0, item 526).

[36] REgular LAnguage for XML Next Generation (RELAX NG) – another XML schema definition language crated by the OASIS group.

[37] This schema was first contested, and later repealed.

and Administration. The final version of the schema can be found on-line in the Polish public registry Electronic Platform of Public Administration Services (ePUAP).[38]

The EDAP schema defines elements for 13 types of legal documents, including, for example, act, regulation, resolution, judgment or announcement. Fig. 13 shows the structure of act (*ustawa*) element.

Fig. 13. The structure of act (*ustawa*) element in Polish legal schema.xsd
Source[39]

The act element contains metadata (*metryka*), (optionally) a reference to the judgment of the Constitutional Court (*trybunal*), comment elements (*elem-komentarza*), elements describing sections of the act: preamble (*preambula*), part (*czesc*), book (*ksiega*), title (*tyt*), division (*dzial*), chapter (*rozdzial*), article (*artykul*), signatures (*podpisy*), approval (*zatwierdzil*), appendices (*zalaczniki*), footnote content (*tresc-przypisow*), binary appendix (*zalacznik_ _bin*), and finally XML digital signature (XMLDSig – *ds:Signature*). The EDAP schema is stored in a single file and it also does not rely on any other

---

[38] http://epuap.gov.pl/wps/PA_1_FCHRSKG108KP5023L0FQM82083/produkt_re-komendacji/produkt_rekomendacji_340 (accessed 28.06.2013).

[39] Visualization prepared in Altova XML Spy (Home) on the basis of EDAP schema.

schemas. In this sense it is similar to the BWB DTD schema. However, in contrast to BWB, it forces the sections of the legal acts to be stored in a strict hierarchy, and allows for the better control of types since it is XML Schema.

Along with the EDAP editor ordered by the Ministry, which was developed in platform-independent and open technologies, like Java and Apache Tomcat, commercial editors were prepared by a private company for the MS Windows operating system. These editors use a schema similar to EDAP, however each element has unique identifiers and some of them may have extra attributes (e.g. "repealed" attribute). Fig. 14 contains a fragment of a regulation (*rozporzadzenie*) with *id* attributes for each element, saved in commercial editor.

```xml
<rozporzadzenie
xmlns="http://www.crd.gov.pl/xml/schematy/edap/2010/01/02"
id="576D21FF-886C-4C9C-BFBC-D171A0F734FD" uchylony="nie"
status="uchwalony" widoczny="tak">
   <!--Legislator 2.1.0.22-->
   <metryka id="DD9848C5-9EDB-4D85-AB98-EC995B4F5AA6" status-
   aktu="uchwalony" nazwa="rozporzadzenie" numer="" opis-
   typu="Rozporządzenie" data="2002-06-20T00:00:00Z" widoczny="tak"
   organ-wydajacy="Prezesa Rady Ministrów" organ-wydajacy-m="Prezes
   Rady Ministrów">
      <organ-wydajacy id="4309074A-8466-44CD-BF7C-A965AA605AFB"
      glowny="tak" w-mianowniku="Prezes Rady Ministrów">Prezesa Rady
      Ministrów</organ-wydajacy>
      <tytul id="25991D03-5B0A-4843-A425-17EAE7420D31">w sprawie
      "Zasad techniki prawodawczej".</tytul>
   </metryka>
   <podstawa-prawna id="F1DB8FC6-6B5C-4EB6-BCA4-F6AD53EC68A5" par="3">
      Na podstawie art. 14 ust. 4 pkt 1 ustawy z dnia 8 sierpnia 1996
      r. o Radzie Ministrów (<link id="AB2C7E1E-CED0-44AF-9663-
      68F484F9648B" typ="hiperlink">Dz. U. z 1999 r. Nr 82, poz.
      929</link>, z<link id="295FA7AD-35DC-4B43-8A8B-9846706A1231"
      typ="hiperlink">2000 r. Nr 120, poz. 1268</link>oraz z<link
      id="74EB4893-34B4-44C1-A86E-63A2DD9E43DA" typ="hiperlink">2001
      r. Nr 102, poz. 1116</link>i Nr<link id="321D17B5-257D-4377-
      A84D-A0E2396B969B" typ="hiperlink">154, poz. 1799</link>i<link
      id="85353E6F-4058-4FD1-A635-71A4BD1CE577"
      typ="hiperlink">1800</link>) zarządza się, co następuje:
   </podstawa-prawna>
   <paragraf id="192E877E-CC68-4457-8F56-C76C626C03C9" nr="1" par="4">
      Ustala się "Zasady techniki prawodawczej", stanowiące załącznik
      do rozporządzenia.</paragraf>
   <paragraf id="80BF31A9-2342-4B59-AF2F-EC285EAD86C2" nr="2" par="5">
      Rozporządzenie wchodzi w życie z dniem 1 sierpnia 2002 r.
   </paragraf>[...]
```

Fig. 14. Internal XML format of legal act in Legislator editor[40]
Source: own elaboration

---

[40] An extract from documentation and related files for the Legislator editor by ABC PRO.

Although the internal format shown in Fig. 12 can be transformed into the form that does not contain *id* attributes, some attributes like *widoczny* (visible) or *data* (date) in the element *metryka* (acts' metadata) cause that the XML document is still not valid according to the original EDAP schema.

The ambiguities and uncertainties associated with the official EDAP schema meant that this schema was repealed by the Regulation of the President of the Council of Ministers of 27 December 2011 on the technical requirements for electronic documents containing acts and other normative legal acts of the gazettes published in electronic form and electronic means of communication and information media (Dz.U. 2011, no. 289, item 1699). According to this regulation "Electronic documents containing normative acts and other acts being subject to announcement […] shall be prepared in XML format". This format should be, in turn, consistent with the templates of acts announced in the central repository for electronic document models (part of ePUAP[41] platform). Templates should contain an XSD Schema that is consistent with the structure defined in the Principles of Legislative Technique.[42] Along with the XSD schema, an XSL file determining how to visualize the act should be provided as a part of the template. All the acts being subject to announcement shall be published in PDF format or in XML format, in the way that allows one to visualize its contents in a PDF document.

4.2. XML schema for on-line consultations

Yet another schema was developed within a pilot project dedicated to the on--line consultation system that was fired in 2010 by the Ministry of Economy. The main goal of the project was to increase the transparency of the legislative process, and to involve a wide range of social groups in this process.

The new schema is organized in a similar manner to the NIR XML schema, in which elements were divided into different files, and analogously it has two main versions: *strict*, designed for the preparation of new acts and regulations, and *loose*, that allows the storage of archive legal acts that may not be fully consistent with all the restrictions existing in the strict version. Other types of legal documents, like resolutions and statements, can be structured according to the *base* version of the schema. All these schemas include common elements like text content (*zawartosc tekstowa*), paragraph content (*zawartosc akapitowa*) that are defined in common (*wspolne*) schema. The relations between schemas are depicted in Fig. 15.

---

[41] Electronic Platform of Public Administration Services (ePUAP), http://epuap.gov.pl (accessed 28.06.2013).

[42] Order of the President of the Council of Ministers of 20 June 2002 on The Principles of Legislative Technique (Dz.U. 2002, no. 100, item 908).
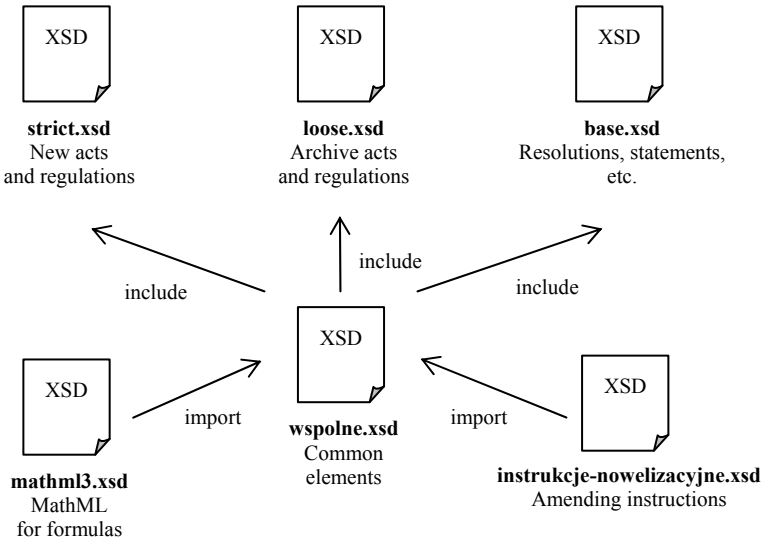
Fig. 15. The organization of schemas in the on-line consultation system
Source[43]

The strict and loose schemas contain 32 elements with two main elements: *ustawa* (act) and *rozporzadzenie* (regulation). Both of these can contain such elements as *data* (date), *tytul* (title), *preambula* (preamble), *zalacznik* (appendix), *artykul* (article – in the case of act), *paragraf* (paragraph – in the case of regulation) and *jednostka-systematyzacyjna* (systemizing, hierarchical unit). As articles or paragraphs are numbered in a continuous way, their content is not placed directly into hierarchical elements (*czesc* – part, *ksiega* – book, *tyt* – title, etc.), but placed on the same level. This is shown in Fig. 16 on the basis of act (*ustawa*) element.

Articles (for acts) and paragraphs (for regulations) can contain enumerations and paragraphs (in editorial sense), that, in turn, can use elements such as text paragraphs (similar to the *p* tag in HTML), tables, raster or vector graphics and equations.
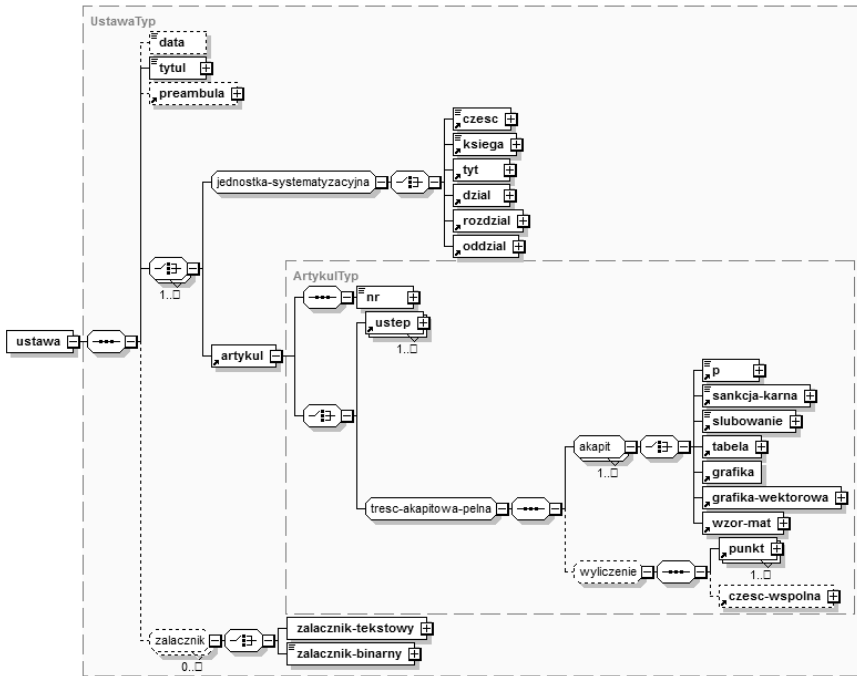

Fig. 16. Act (*ustawa*) element in the schema used in the on-line consultation system
Source[44]

4.3. Polish standards and transnational legal schemas

Further integration within the European Union cannot proceed in isolation from the integration of the legal systems of its Member States. The first step towards such integration should be in the form of a common database of all legal acts that are in force in individual countries. In the majority of these countries, legal acts are currently promulgated in electronic form and, in many cases, electronic documents containing legal acts are treated as authentic and have legal effect (in some countries, including Poland, since January 1st, 2012, they are the only source of new law[45]). Although the European system called N-Lex[46] is able to query national databases according to the

---

[44] Visualization prepared in Altova XML Spy on the basis of *Dokumentacja techniczna schematów XML Schema...*

[45] As on November 1st, 2011, on-line version of legal gazette are exclusively authentic in: Belgium, Denmark, Spain, Hungary, Netherlands, Austria, Portugal, Slovenia (also in non-EU members: Iceland and Norway), source: Online publication of legal gazettes in Europe (EU/EFTA), http://circa.europa.eu/ (accessed 28.06.2013).

[46] N-Lex was launched on 28 April 2006 and initially covered only 4 countries.

words and value of parameters like the type of act, date of publishing, that are provided by a user, it depends on different API[47] of national legal information systems. As a result, a user cannot always pass all the parameters – e.g. in the Portuguese system, terms to be searched for cannot be inputted, only the number of legal act while for the Czech legal system neither the type of legal act, nor publication date can be provided to the query. Moreover, if the national system does not respond in a given time, the results of the query cannot be displayed. Undoubtedly, a lot more could be achieved if all texts of legal documents could be gathered in one system in the form of XML files that follow a common schema. This would unify and speed up the search process itself, but, first of all, would enable cross searching in legal texts of all EU members, and, in the future, application of different knowledge processing tools which could further facilitate access to legal texts for citizens.

Currently, there is no single standard for legal documents description that is recommended by the European Commission, however, the two standards described in the previous section seem to be the best candidates for such a common European legal standard: Metalex that is currently developed by CEN Committee, and Akoma Ntoso that has been recently adopted for AT4AM editor, used at the European Parliament.[48]

As mentioned before, in the project initiated by Hoekstra[49] on the translation process of Dutch law written in BWB format into Metalex standard, the translation between these two standards in most cases could be done automatically, as all BWB documents follow the same structure. If legal documents do not follow the same schema, the whole process can be much more complex and, as a result, can be affected by many more translation errors.

The document describing the schema developed for the on-line consultation system for the Ministry of Economy[50] contains a short analysis of the conversion capabilities from this schema into Metalex and Akoma Ntoso standards. In the authors' opinion, the most troublesome aspect in the translation process can be vector graphics written in SVG language, equations written in MathML, references for raster graphics and binary appendices saved in external files and, finally, tables (Metalex does not have any dedicated elements for storing tables), and quotations. Similar problems should not affect the conversion into Akoma Ntoso standard.

---

[47] Application Programming Interface (API) – the set of rules allowing for interaction with a given system.

[48] AT4AM for ALL, the web editor for amending legislation of the European Parliament, is released in Open Source, April 1, 2013, http://www.akomantoso.org/rss-manager/at4am-for-all-the-web-editor-for-amending-legislation-of-the-european-parliament-is-released-in-open-source (accessed 28.06.2013).

[49] R. Hoekstra, *The MetaLex Document Server – Legal Documents as Versioned Linked Data...*

[50] *Dokumentacja techniczna schematów XML Schema...*

The EDAP schema does not contain elements dedicated for graphics or equations, however, it enables for so-called islands of data, i.e. text fragments that can be placed outside the regular structure of an legal act. This can cause potential problems with translation into Metalex or Akoma Ntoso schema. Other potential problems may result from the definition of events, like the repealing or modification of an act. Both Metalex and Akoma Ntoso allow for the definition of such events in special elements, however, the philosophy behind their construction is different. Fig. 17 indicates, on the basis of preamble element, that only a few elements and attributes can be directly mapped from the EDAP schema to Akoma Ntoso.
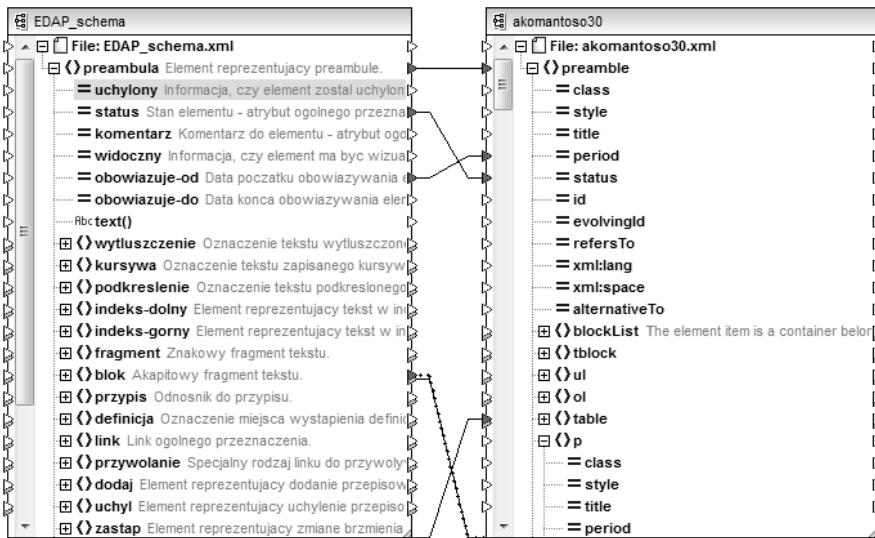

Fig. 17. Preamble element mapped from EDAP schema into Akoma Ntoso schema
Source[51]

Formatting elements that in an EDAP schema can be directly placed under a root element like bold (*wytluszczenie*), italic (*kursywa*), and so on, have to be put in some HTML block element in Akoma Ntoso (e.g. paragraph, list item, and so on). Elements like add (*dodaj*), repeal (*uchyl*) or replace (*zastap*) have to be defined in event or textual modification elements. These and similar difficulties, however, cannot be overcome unless all legal documents strictly follow EDAP or another common schema. If legal documents

---

[51] Visualization prepared in Altova Map Force on the basis of EDAP and Akoma Ntoso 3.0 XML Schemas.

are defined in XML files with different structure, each type will require an individual definition of conversion procedure and this can be very difficult in practice.

5. Conclusions

In recent years, legal acts and other sources of law have undergone a process of rapid transformation from their traditional paper form to that of electronic documents. The rapid increase in the percentage of citizens having access to the Internet has created the need for such initiatives as freedom of information laws (FOI), as reflected by, among others, in the Electronic Freedom of Information Act Amendments in United States (1996), or in the directive on the re-use of public sector information in the European Union (2003). Legal acts written in an electronic form have steadily replaced the traditional printed form, in many countries becoming the only authentic source of law.

On the other hand, the integration process in Europe and globalization of world economy require instant access to not only national legal documents, but frequently multinational sources of law. Whether it is possible depends on a large extent on the form in which legal documents are prepared and published. The XML language is currently the common, universal format for exchanging information over the Internet network, and allows for the semantic description of legal document content. Although many national and transnational XML standards have been created, there is no single, widely acceptable standard that can be used in the whole legislative process, from legal drafting up to promulgation of the approved act. This is true not only for multinational organizations, like the European Union, but also for individual countries, like Poland. The positive examples of common national standards developed in many countries (e.g. Italy, Denmark, the Netherlands, Switzerland, Brazil or New Zealand), and a multinational standard for African parliaments (Akoma Ntoso) leads one to believe that the development (or adoption) of such standards, and their wide acceptance can be achieved in the near future. This will certainly allow for an entirely new quality to emerge in terms of access to legal information by citizens and the automatic processing of such information.