Medical Informatics Europe 2002
G Surján et al. (eds)
IOS Press 2002

The role of XML in medical informatics in Hungary

Dr. Nándor BALOGH

Semmelweis University, National Health Insurance Inst. Budapest
balogh.n@emk.sote.hu balogh.n@oep.hu

Abstract. Reforms in healthcare means frequent changes. The author focusing on flexible information systems and data structures in healthcare, that gives the chance to avoid costly deadlocks, incompatibilities and data chaos -keeping the financial and administrative burden of provider and reimbursement information systems at low level.
He describes data and document exchange among health care providers (hospitals, GPs etc.) and the reimbursing insurance institute introducing XML (eXtensible Markup Language) technology and it's multimedia extension called SMIL (Synchronized Multimedia Integration Language) that well supports large textual and multimedia nature of data in healthcare.
He presents the most frequent XML datasets for reimbursement in Hungary, the validation schemas (Data Type Definitions) and procedures, comparing the new and old data file sizes with different compressions. The comparison includes different schema file sizes, and validation times.
Keywords: teleinformatics, telemedicine, XML, SMIL

1. Methods and results

XML was developed for data exchange among heterogenous textual and relational databases. This includes the most common problem in medical informatics that exists inside hospital (discharge reports and tests), between GPs and hospitals and between reimbursing institute and the suppliers. XML was developed to exchange self-describing datasets simply among different databases.

In XML data fields are separated by tags that are put between < / > characters. The markup syntax of XML is: **<tag>** Datafield **</tag>**

XML focuses on data structure rather than on presentation. There are no predefined tags, so it is very flexible like:

<Sirname>Balogh**</Sirname>**
<Name>Nándor**</Name>**
<InsuranceNo>12345678901**</InsuranceNo>**

This (UTF-8 coded by default) text file is saved as data using .xml extension. This means the data to be exchanged can be read and fully understand by human.

Tags can be defined by either data exchanging party. This might be done totally free or by some administrative or professional body in the form of **Data Type Description (DTD)**, XML Data, W3C Schema, BizTalk Schema etc. In either case this definition can be described (fortunate case in XML).

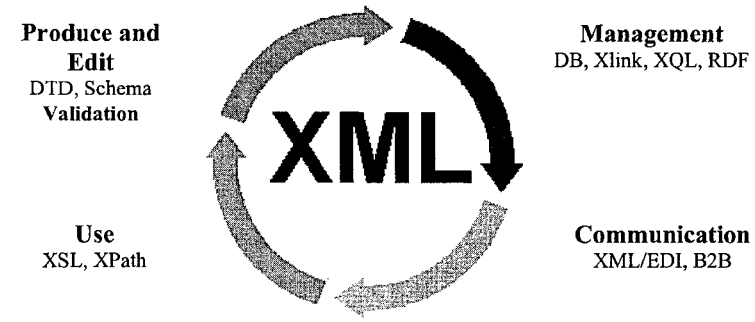


Figure 1: The different parts of XML technologies

We can compare the XML data file to the DTD. If it compares the XML is valid, otherwise (if the syntax is OK) it is only well formed.
XML focusing on pure data, it doesn't deal with presentation. An XML file can be presented in different ways. For intelligent presentation we use style sheets and lately XSL (eXtensible Stylsheet Language).

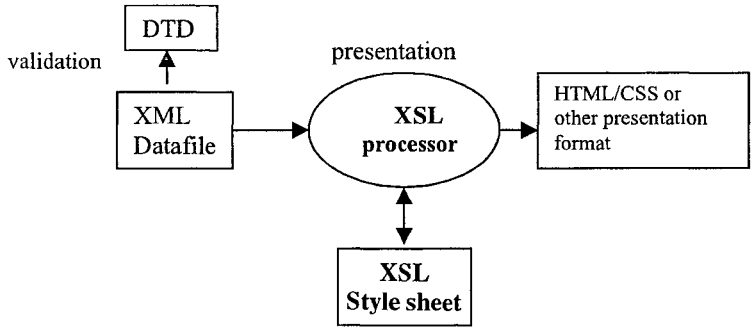


Figure 2: The XML validation and presentation process flowchart

Latest Web browsers contain XML parser and some RDBMS also supporting XML. There are many XML editors, parsers, validation and compression, presentation tools and converters.

The demonstrating datasets:

XML dataset:

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- edited with XML Spy v3.0.7 NT (http://www.xmlspy.com) by BN (OEP) -->
<!DOCTYPE BIZTOSÍTÁS SYSTEM "E:\xml\doc\Hattan\BNdemo.dtd">
<?xml-stylesheet type="text/xsl" href="E:\xml\doc\Hattan\c.xsl"?>
<BIZTOSÍTÁS>
        <Páciens>
                <Vezetéknév>Balogh</Vezetéknév>
                <Keresztnév>Nándor</Keresztnév>
                <TAJszám>12311238901</TAJszám>
        </Páciens>
        <Páciens>
                <Vezetéknév>Kiss</Vezetéknév>
                <Keresztnév>Zsolt</Keresztnév>
                <TAJszám>14302258912</TAJszám>
        </Páciens>
</BIZTOSÍTÁS>
```

DTD dataset:

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- edited with XML Spy v3.0.7 NT (http://www.xmlspy.com) by BN (OEP) -->
<!--DTD generated by XML Spy v3.0.7 NT (http://www.xmlspy.com)-->
<!ELEMENT BIZTOSÍTÁS (Páciens+)>
<!ELEMENT Keresztnév (#PCDATA)>
<!ELEMENT Páciens (Vezetéknév, Keresztnév, TAJszám)>
<!ELEMENT TAJszám (#PCDATA)>
<!ELEMENT Vezetéknév (#PCDATA)>
```

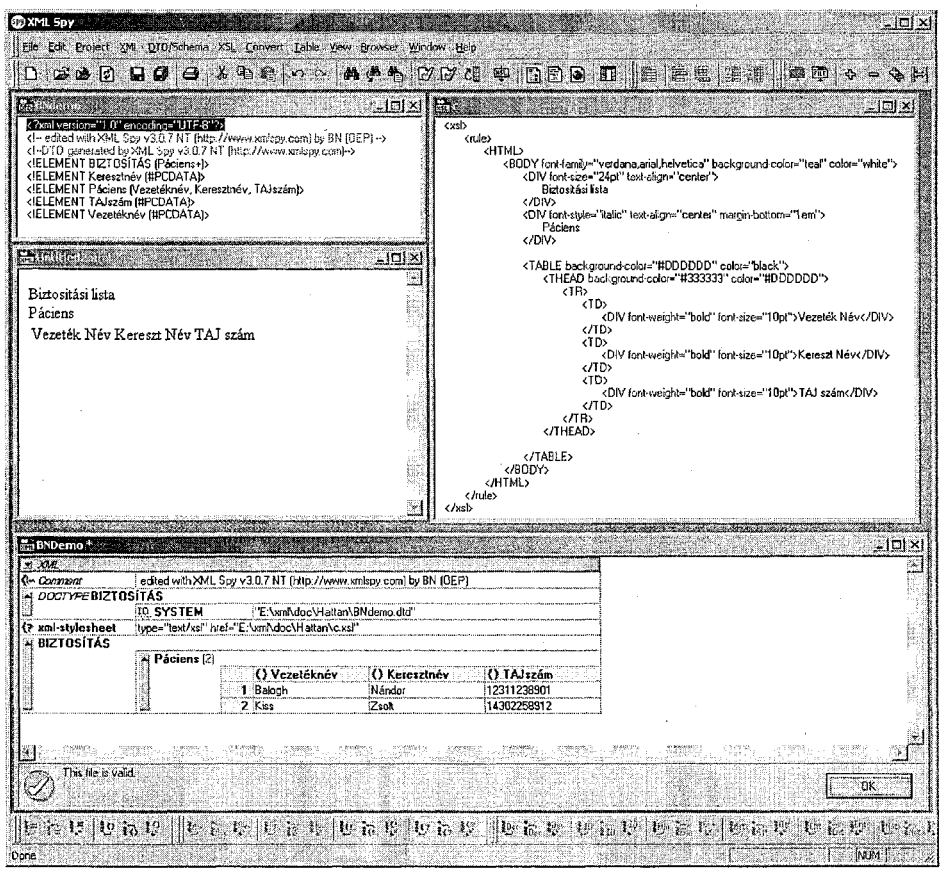


Figure 3: The XML validating process screenshot (The file is valid)

2. XML among health provider institutions

As patient moves among health institutions his/her data should move too. Test requests and results, prescribings, discharge reports should be exchanged on-line.

Document (eg. Discharge report) and data exchange is not simple even inside hospital departments as the reports contain data fields from RDBMS-es (like lab data).

The problem I solved at the beginning of the 90 ties very similar to the markup language used today (using $\leq$ és $\geq$ markups). From the database the maker of the report could select those parts he like to include in reports that is put between the above tags on screen, but it will not be printed.

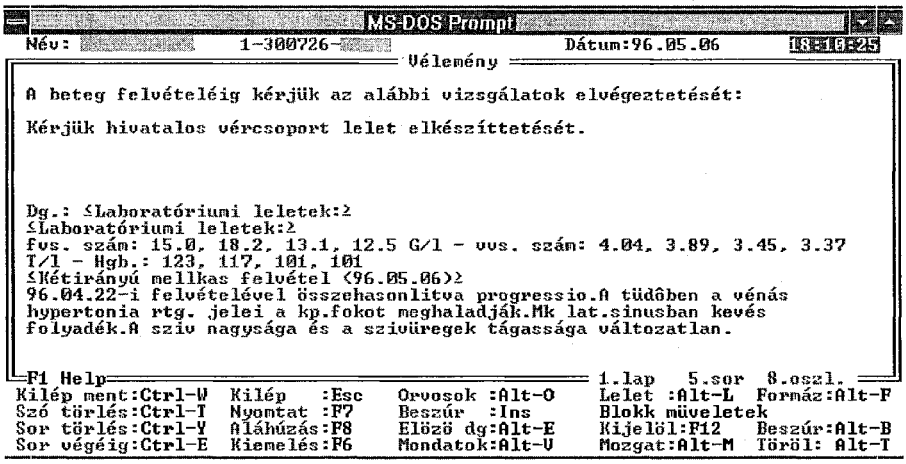


Figure 4: XML like tags used for our Discharge reports at Hungarian Institute of Cardiology

Few years later XML found the solution that accepted and supported by the industry (we should program everithing ourself).

At the end of 90s in the framework of the European SAMTA telemedicine project we used XML for angio and ultrasound reports.

The ultrasound look like this:

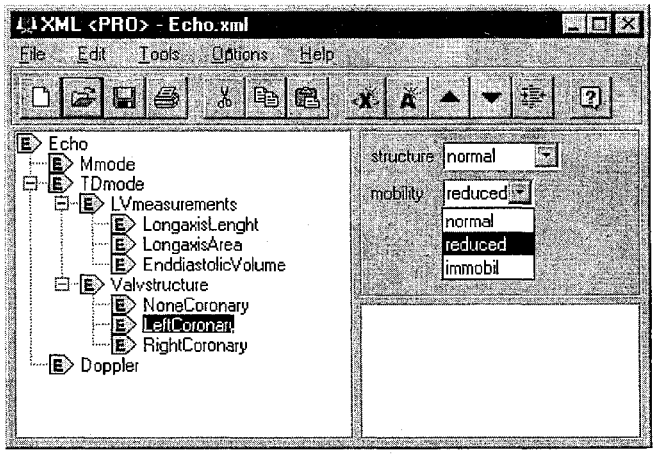


Figure 5: Part of Ultrasound test result XML file

SMIL is an extension of XML to handle sound, text, image and movie information to play in predetermined sequence or parallel. An example I've used during our SAMTA project with angi and ultrasound image, loops, text and sound:

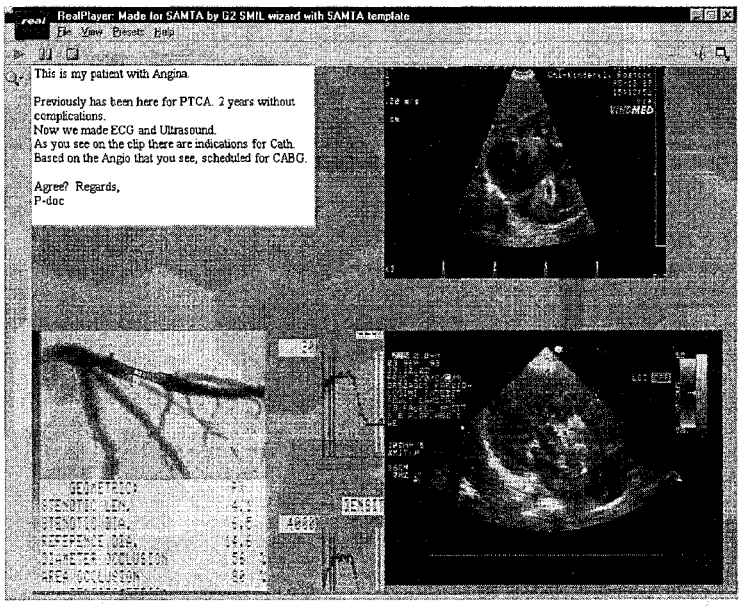
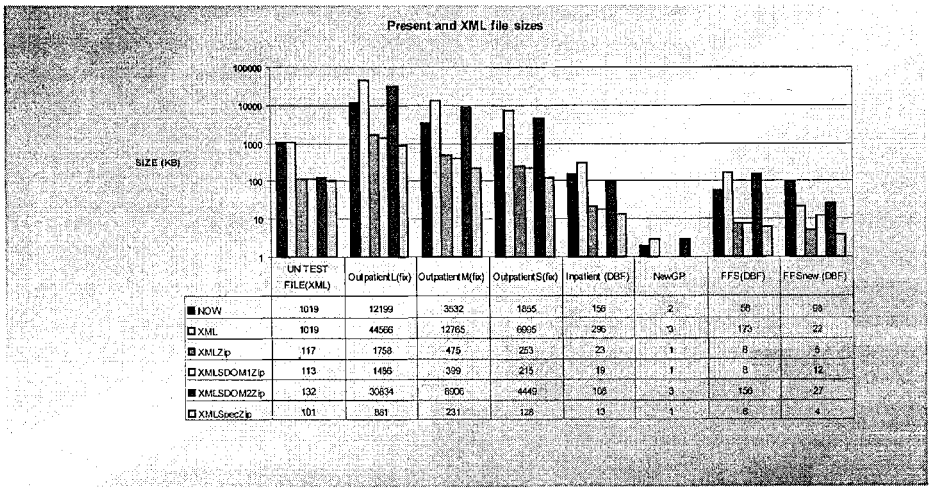


Figure 6: SMIL screenshots gor cardiac angio and ultrasound muntimedia data

3. XML in data exchange between health providers and insurance institution

In Hungary we just reengineering and change the information system for reimbursement at National Health Insurance Institute (NHII). I've examined how the heterogenous formats of present exchange formats (fix lenght records, DBF, Excel etc.) can be transferred to XML and how tey can be validated by providers to a DTD provided by NHII. The most critical the outpatient data produced by the University clinics. This data is sent in one package in fix record lenght format and can be as large as 12 Mbytes as seen below.



	UNTEST FILE(XML)	OutpatientL(fix)	OutpatientM(fix)	OutpatientS(fix)	Inpatient (DBF)	NewGP	FFS(DBF)	FFSnew (DBF)
NOW	1019	12199	3532	1895	156	2	58	58
XML	1019	44566	12765	6905	296	3	175	22
XMLZip	117	1758	475	253	23	1	8	5
XMLSDOM1Zip	113	1456	399	215	19	1	8	12
XMLSDOM2Zip	132	30834	8606	4449	106	3	156	27
XMLSpec2Zip	101	881	231	128	13	1	6	4

Figure 7: Different XML compression for different healthcare provider types in Hungary

As can be seen the size of XML is larger than present used ones (12MB grown to 44 MB). We can compress XML however quite well, depending on the content. For the outpatient data, the best result is given by algorithms using the speciality of XML tag syntax. This is followed by DOM1, WinZip and DOM2 respectively. As the original data files are sent usually in compressed way, so we need to compare the compressed data sets. Theoretically DOM compression is better, but there are practical issues (like market share and longevity etc.) as well. Based on the compression above we can use XML even the original XML file sizes are larger (a 44Mbyte XML became 881Kbyte using XMLspeciZip).

After (sometimes problematic) installation of the compression software packages, the time required for compression is not substantial, it was below 1 minutes even for the largest datasets.

Size of XML dataset	Load	**VALIDATION TIMES (SEC) AND SCHEMA FILE SIZES (KBYTE)**								Replace	Conversion
		DTD	**size**	**Biz Talk**	**size**	**W3C**	**size**	**XML Data**	**size**		
4000 records	120 s	30 s	2KB	50 sec	6KB	25 sec	4KB	40 sec	6KB	180 s	60 s
65527 records	3 h	too	long	wait	ing for	menu	Items			2 hs	2 hs

Figure 8: Times required for XML validation and other processes for different XML data sizes

From the above we can see that the validation times at client side are too long (for 65 th records). Validation however can be done at server site or separately by the different departments of a large outpatient clinic (to get the above 4 th records).

4. Conclusions

XML and SMIL have large TECHNICAL potential for data exchange among healthcare and reimbursing institutions. The difficulty is the CONTENT represented by DTD and tag definitions. We can use SNOMED or National Library of Medicine (USA) definition of terms to define Hungarian tags. This job can be done in the framework of different medical societies and advisory bodies using the appropriate motivation for their members.

References

[1] N. Balogh: European Project SAMTA, Deliverable 3, 11
[2] N. Balogh: XML in helathcare. XXII Meeting of the John von Neumann society, Veszprém, 2001