eXtensible Markup Language: Its Application in Development

A report on the application of XML in developing countries

RESEARCH REPORT
No. 1, January 2001

eXtensible Markup Language: Its Application in Development

A report on the application of XML in developing countries

A research study by ITM Management & Informatics for the International Institute for Communication and Development

Jaco Jellema
Jaapjan Verboom

Content

INTRODUCTION

The overall objective of the research was to gain insight into the relevance of XML for IICD's local partners in developing countries. The following objectives were formulated:

- To give an overview of interesting (for IICD's local partners) information resources about XML.
- To understand the transformation process from existing data into XML/IDML-formatted data.
- Understanding the required processes.
- Understanding the technical consequences.

ITM Management and Informatics (http://www.itm.nl/) did the research and documentation which was presented in both electronic form and hardcopy.

Structured Information

The use of information technology to support internal organisational processes is becoming the norm. Automated processes require standardized information streams and structures such as databases.

The kind of information published on the Internet in the first phase was in general not structured for processing by computers, but for presentation directly to the end user. Corporate websites, for example, were the digital frontdoor where an organisation presented itself. Commercial websites are increasingly becoming the ´backdoor´ to the back-office where the real business happens. The change from presentation to transaction oriented information exchange is possible when the processes and information are well structured. eXtensible Markup Language (XML) is one of the best ways in which structured information can be exchanged.

Markup Technology

The World Wide Web was a revolution for the Internet. The next step in the revolution will be eXtensible Markup Language (XML). The Internet protocol connects computers to each other, which allows e-mail and transferring of files. On the World Wide Web, by using Hypertext Transfer Markup Language (HTML) it was possible to link documents and view them in a browser. This made information easily accessible to Internet users. HTML uses tags to indicate formatting and attributes of text, this is called markup. For example "<H1>" indicates the start of the heading that would be text in bigger letters or emphasis compared to plain text.

In XML tags no longer indicate formatting of the documents, but instead indicates elements or fields of structured data. The interpretation of the tag now depends on the application that uses a Document Type Description (DTD) to interpret and check the data. The DTD, or schema, describes the possible fields in an XML document. The formatting power of HTML is retained and increased by the separation into style sheets that can be used to create different formats of presenting and transforming the data. The XML revolution is the emphasis on the context of text; a step further than only formatting. The XML documents are structured to be machine-readable and therefore to support computer interaction such as Electronic Document Interchange (EDI).

The example below is a fragment of an XML document based on the International Development Markup Language Document Type Declaration, describing a project funded by IICD. Each tag has a corresponding end tag with the same name, which starts with a slash that indicates the end of a section. Some tags contain an attribute such as xml:lang="en" which in this case indicates the language of the field as English. The order of the fields is based on the DTD that also defines the attributes and whether a field is required, optional or repeating.

```
    <activity>
        <id_part>
            <activity_id>IICD_1502.01</activity_id>
            <title xml:lang="en">Business Intelligence Trade Points</title>
            <funding_org>
                <org.acronym>IICD</org.acronym>
            </funding_org>
        </id_part>
        <desc_part>
....
        </desc_part>
...
    </activity>
```

In general, users do not see this source code, but stylesheets are used to present this in another format. The eXtensible Stylesheet Language (XSL) is used for formatting, while XSL Transformation (XSLT) can change one XML document into another XML document or even a different file format. This means that one XML document can undergo various transformations to, for example, HTML or pdf.

The following resources give more detailed information on XML:

- Investing in Knowledge Toolkit (*http://www.oneworld.org/thinktank/iktools/*)
- XML in 10 points *(http://www.w3.org/XML/1999/XML-in-10-points)*
- BUILDER.COM - Web Authoring *(http://www.builder.com/Authoring/Xml20/ss05.html)*
- FAQ with answers *(http://www.ucc.ie/xml/)*

Research Overview

In this chapter a short introduction on the research and XML was given. In the second chapter the relevance for developing countries is discussed. The third chapter goes into XML tools that can be used to edit XML documents and how these can be compared. Finally, the fourth chapter summarizes the conclusions.

XML IN DEVELOPING COUNTRIES

The application of XML always occurs in specific sectors where organisations work with the same type of information. In this section two cases are discussed. The second section gives an idea of what types of XML applications can be relevant within a sector. Finally, the last section sketches the dynamics of the development of XML related technologies that can be seen as a source of opportunities for organisations in developing countries.

Cases

The software sector is of course a first-mover in the take up of new information technology. The first case shows how marketing of products in various ways can be facilitated using XML. The second case is in the international development sector, which is a sector with extensive inter-organisational information exchange.

Portable Application Data (PAD)

A small software company in Uganda develops software packages for the international market. They regularly release new updates and versions, which are distributed using the Internet. There are 30-day evaluation versions available free of charge from major software sites and from the company's own web site. With each release the company's web site needs updating, but also the international resellers need to be informed. This requires accessing their websites and updating the information, using forms on the web site. This needs to be done for each site that refers to their software.

The Association of Shareware Professionals (ASP) has recognised the problem for small software developers all of the world. They have defined a standard description of software, which is called Portable Application Data (PAD). The information is stored in an XML file that can be processed in a few different ways. The files can be created with a free programme that guides the user in the creation of a PAD file. This file, when made available on the company's web site, can be checked regularly by data aggregators who sell software packages from different producers. The company can also use the file to produce different kinds of summaries for their web site and email or hardcopy distribution.

The Ugandan software company can choose to use the data structure proposed by the Association of Shareware Professionals. This can be done by using tools offered by them or by using general XML tools. This should ease marketing and updating of the company products.

http://www.asp-shareware.org/

The advantages summarised:

- Sharing of best practices (in software description);
- Supporting structured low-volume data production;
- Supporting aggregation of data;
- Producing up-to-date output in multiple formats.

International Development Markup Language (IDML)

Within the international development community there has been an effort to exchange project information to enhance aid effectiveness. At a country level United Nations Development Programme collects information about aid activities and other organisations, such as the Organisation for Economic Co-operation and Development, attempt to do so at a global level. The International Network for Development Information Exchange (INDIX) supported these initiatives by developing Common Exchange Formats for Development Activity Information (CEFDA) and also by distributing CD-ROMs with the collected information.

The International Development Markup Language Initiative (IDML initiative) builds on the existing effort that was started before the age of Internet. The CEFDA standard was used to create a DTD for IDML. There have been a few pilot projects to test the possibilities of searching through IDML data that is stored at different sources. The organisations in general produce the IDML data dynamically from existing information systems and offer it on their own web site. The strict use of XML will mean supporting the Unicode character set, which makes it possible to handle non-Latin character sets.

http://www.idmlinitiative.org/

The advantages summarised:

- Organisations can keep control of their own data;
- Powerful searching and aggregation of data is possible;
- Multilingual data contribution is possible.

Ten Best Bets

To get an idea of the relevance of XML for organisations in developing countries, an analysis was done of the article `Ten best bets for XML applications´ (Schloss 2000). The article considers applications that are especially suited for the use of XML. Each of these applications is based on aspects of XML, which were then considered in the context of developing countries.

1. Applications that use data with repeating text fields

Applications whose persisting data includes a variable number of repeating (nested) textual fields should use XML. There are advantages in the storage and accessibility of the data. This is a general advantage of XML and not specifically relevant for developing countries.

2. Applications that need Unicode

Applications for multilingual or multinational audience that use textual names of people, enterprises, and geographic places need the flexibility to display the names in the correct local language and character set. Use XML for such applications because XML uses Unicode by default, which allows end users to encode names in their preferred language and character set.

This aspect of XML is expected to increase the attention for multilingual applications, even if this only concerns the Latin character set. This is important for developing countries where often even within one country many languages are used.

3. Applications that exchange data between enterprises

Another obvious use for data exchange formatting is an application that exchanges data between enterprises, using software produced and configured by more than one enterprise. The benefit of XML here is debugging, because the data interchange is at least somewhat readable when done with XML tags. The advantage of debugging is not specifically relevant for

developing countries, but the advantages in general of data exchange will be discussed in point 7 and 10.

4. Route-and-flow applications

Route-and-flow applications differ from conventional applications: they do not have one programme active for an entire transaction, waiting in real-time for subsystems to complete their portion of the work. Instead they use independent subsystems that examine or augment the data and then pass it on to one or more other subsystems, which can work asynchronously, using some type of queued data pipe. XML is particularly suitable right now for those route-and-flow applications that do not have real-time transaction confirmation requirements.

In developing countries the route-and-flow applications can ensure robust data exchange over national and international networks. The independent subsystems can communicate over low bandwidth networks that do not need to be very reliable.

5. Applications that produce microcode data reports

XML's structured repetition and programming language independence means new generations of server codes can evolve to collect data from embedded sensors without changing the embedded-device microcode. This does not seem especially relevant in developing countries.

6. Applications that display results on various devices

As ubiquitous computing takes hold, more and more applications must present data to end users on many varied devices. This is a natural use of XML: the applications present status to end-users as an XML document. Transcoding would alter the data for optimal rendering on the end user's current access device, whether handheld computer, Web browser, pager, audio phone, smart cellular phone, or other. In developing countries ubiquitous computing or being continually available for communications is not directly an issue. What is relevant is the flexibility for different ways of displaying data to end-users. Mobile phone rendering involves compact and concise information exchange that can also be relevant in other low bandwidth environments.

7. EDI-like or supply-chain applications

Although standards in this area are still evolving, applications that offer functions that work like EDI (electronic data interchange) are pretty good bets for XML development. Most of the value of EDI can be duplicated using XML, extranets, and Virtual Private Networks, as well as the emerging IETF/W3C standard for Digitally Signed XML documents using Public Key Infrastructure encryption keys.

Pairwise supply chain applications, that perform the interchange via the Internet rather than a Value Added Network (VAN) provider, are the best candidates to begin developing now. XML over the Internet therefore becomes a facilitator for international Business-to-Business e-commerce for companies in developing countries. Building on the EDI experiences, importing and exporting companies may find it easier to participate in international marketplaces.

8. Applications that need enterprise document management

Documents, unfortunately, don't always contain the full context to allow them to be properly interpreted. So enterprise document management must be more than just databases or digital libraries; it must contain *metadata* -- data about the documents and about classes of users, classes of operations and applications, individual users, and more.

XML works very well for the metadata that allows versioning, combining in numerous configurations, user-interest profiles, and so forth. This is relevant for large enterprises but also for institutions managing large information flows. Sectoral standards for the metadata do not need to be specific for developing countries and therefore the sharing of these standards can offer much value.

9. Enterprise application integration

Crafting applications for integration of Enterprise Resource Planning (ERP) systems, transaction systems, repositories, and groupware is now possible. Connectors and adapters that allow these systems to accept input and produce output in XML formats are quickly coming into the market. Mismatches between the outputs of one system and the inputs of another can sometimes be fixed with a simple transformation rule expressed in XSLT.

These issues are not specific for developing countries and depend on the uptake of XML by vendors of applications such as a database. The flexibility and potential in environments with complex information systems and interactions is clear.

10. Matchmaking marketplaces

If you operate a brokerage of some sort, where attributes of available supply sources and attributes of current and projected demand are specified using multiple dimensions, XML stands out as a natural match. Suppliers describe what they have in XML, buyers describe what they are looking for, and your application logic optimizes the matches. In many matchmaking marketplaces, controlled disclosure is required to encourage buyers and sellers. XML easily offers the ability to distribute subsets of supply descriptions, produced with XSLT rules that include only the information that potential buyers have the right to see.

Disclosure policy is not only relevant in marketplaces but is a big issue in the relations between governments and bi- or multi-lateral donors. For documents and data stored in XML these policies can be formulated in transformation rules. This can enhance the transparency while ensuring the autonomy remains.

The exercise of translating the XML applications to developing countries shows that some of the expected advantages are not directly relevant. For example the micro code data reports from embedded systems (5). The three main areas of interest are the multilingual aspects (2), the route-and-flow applications (4) and multi-device displaying (6). Especially for the private sector there is a potential in international (7) and local market places (10).

XML Related Technologies

The development of XML started in 1996 but is in fact based on the Structured General Markup Language (SGML) that was developed in the 1980s. XML has been a W3C standard since 1998, but related technologies are still under development. XSL, the style sheet language, was standardised at a later stage. Interesting developments are the metadata frameworks for XML such as the Resource Description Framework (RDF). It integrates a variety of Web based metadata activities such as site maps, search engine data collection and distributed authoring using XML as interchange syntax. Another development is Xlink, which is a candidate recommendation that provides a framework for creating more complex links than is possible with HTML. Instead of only one-way links, bi-directional and links with multiple destinations will be possible. This is just to mention a few developments that will take time to finalise and be integrated in applications. The using of XML will give experience and knowledge, which can support the evaluation of further developments.

TOOLS FOR EDITING XML

XML is a relatively new standard with many tools still being developed. Further standards to complement XML are also being developed, so this is a very dynamic market. In this section an overview of different types of XML tools is given. The process for the comparison of tools is described. Finally the results of the comparison of XML editors are presented.

XML Software

The following diagramme shows the relationship between different kinds of XML tools when creating a XML document. Sometimes these are included in one application, but one can recognise modules that have similar features. The process shown is a simple publication on only part of an interactive communication process.

The modules in the diagramme are described in the following section.

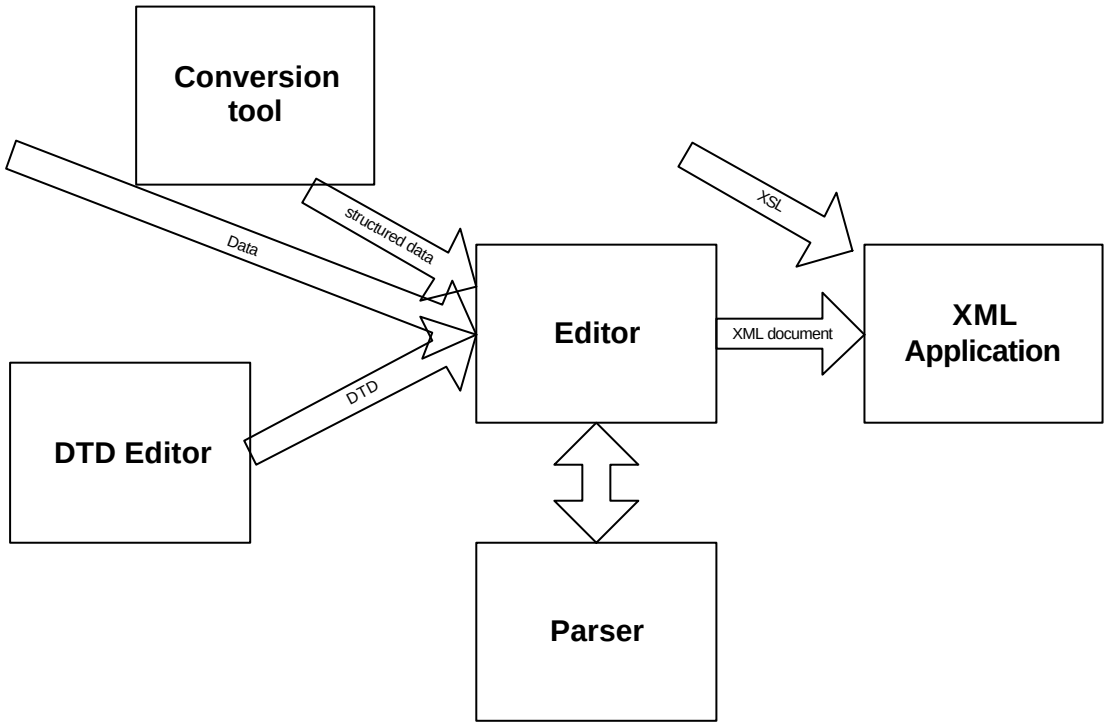

Diagramme: XML software modules

XML Editor

An editor produces XML documents either from scratch or by changing existing documents. Many editors include functions that are described below as separate modules. An editor can have various ways of showing the data such as tree mode, source code mode or preview mode.

The tree mode uses the DTD to show a structured view of the data. In source code mode, one can see the actual XML tags and data. The preview mode is actually an interface to a browser

for viewing the XML documents. An XML editor should ensure that the XML code is 'well formed', which means it should comply with the specifications of XML. Another check of the data is whether it is valid according to the DTD, but this is done by the parser, which is a separate programme.

XML Application

XML is a machine-readable document format that can be processed in many different types of applications. Some XML documents can be presented to users using a browser. The browser will use the XML file and also a template, such as a Cascade Style Sheet (CSS/XSL), which dictates the fonts etc. in the viewer.

DTD/Schema Editor

The Document Type Declaration (DTD) describes the data structure of a XML document. This is done in a different language to XML, which was developed for SGML. Schemas are similar to DTD's but are described in XML, it is a newer standard and not supported by all DTD editors. DTD editors can change the structure such as is described in a DTD. Some editors can propose a DTD based on an existing XML document. Extra features can be the graphical representation of a DTD.

Parser

A parser compares an XML document with a DTD to check if the structure is correct. For example the DTD may indicate that certain data fields are compulsory, so for the XML document to be valid these fields must be included. All applications that process XML documents need to check if the document is valid. For this they will use a parser.

Conversion Tool

The conversion of data to XML is possible only if the source has a structure that can be mapped to the DTD. Conversion tools are usually either an import module of an XML editor or export module of a spreadsheet or database programme.

Selection process

The process of searching for and comparing tools is described so that it is repeatable at a later stage or for users who require different functions. The process includes the finding of tools and the stages of selection and comparison. The long list is the summary of all the tools found during the process. A first selection is made to create the short list, this is then compared on the basis of the criteria. In this section these steps are described.

Long List Creation

To gain an overview of available XML editors an Internet search was done for meta-lists that refer to various editors. An attempt was made to obtain different types of lists, such as informative sites, XML users, and XML researchers. These lists were combined to create a long list of available tools.

Short List Tools

The long list was analysed for duplicate entries; those editors mentioned on more than one list. All these editors with more than one reference were selected for the short list. The packages that did not have an evaluation version available where also removed from the short list. The products which were removed in this process were checked on the basis of the available description to make sure they would be included if there was significant reason to do so.

Evaluation Criteria

The evaluation criteria were used to compare the short listed tools. These were divided into three types: Formal criteria, editor functions and extra features. The evaluation based on the formal criteria gives a basic description of the software package. The second group of criteria indicates which functions are available specifically for editing XML. The final category is extra features, which show how the editor interacts with other modules (see diagramme).

Formal Criteria

The formal criteria describe the origin of the software. These are:

- Name of the product
- Name of the software-developing organisation
- Web site address for information and download
- Version
- Estimated release date (based on timestamps of files but preferably published in release notes)
- Cost and the number of days available for evaluation
- Platform supported (all editors were tested under windows 98)
- What requirements are there for installation (Microsoft XML parser, Java or other)

Functionality

- Is it possible to load (assign) a DTD
- Does the editor support source code editing
- Is it possible to edit in a structured view (for example tree based)
- Is the creation of an XML document based on a DTD easy
- Is there an interface to the parser for validation
- Are help and sample files available

Extra Features

- DTD editor: Is it possible to edit DTD's
- Conversion tool: Is importing of structured data from databases or spreadsheets possible
- Parser: Is there an interface to the parser for validation
- Application interface: Is there an interface to a browser for previewing the documents
- General comments (including extra information such as stability and extra features)

Tools Comparison

The results of the comparison are presented in the following sections. In the first sections the long list of tools is created based on Internet resources. In the second section the results of the selection process are presented. The third section goes into some detail as the tools are compared according to the criteria. Finally some practical experiences with the editors and conclusions have been documented.

Identification of Lists

The following sources of references to XML editors were identified:

- Investing In Knowledge Toolkit -- the results from an IICD sponsored research into XML for developing countries. This was done in the first half of 2000.

- The Data Documentation Initiative -- a project for the documentation in XML of metadata. The format is being considered by a department of the World Bank which gathers data on household surveys done in Africa. The software list is part of the final report which includes evaluation of software for marking up data to XML/SGML.

- Academic XML users in the Netherlands – several meetings every year, parallel to an on-line discussion. One of these meetings (April 2000) was a demonstration of XML editors.

- The Open Directory Project has a list of XML resources and specifically XML editors.

The lists of editors are clearly of different types; it would also be possible to select lists from only category IV. The advantage of different types of lists is that editors are included on the lists for different types of reasons and from different perspectives. For example, the editors mentioned in list III have proven to be useful to academic users. In contrast, list IV attempts to give an overview of options.

Short Listing of Tools

The four lists identified gave a long list of 56 editors (see appendix) of which 22 where mentioned in only one of the lists. When these and all duplicate entries were removed this resulted in a short list of 12 editors.

The following five products on the short list were not tested. For Adept, Clip and WordPerfect this was because no evaluation version was available. Emacs and Xeena did not work in the Windows 98 test environment.

1. Adept

   This product has been integrated with Epic Editor, which is a SGML/XML publishing tool for which no evaluation version was available. The original product evaluated by the data documentation initiative was expensive, this is also true for the current product where the simplest version is over 500 US$.

2. Clip

This Java based editor was not available from the Internet sites indicated by the lists at the time of testing.

3. WordPerfect (http://www.corel.com/Technical_Marketing_Documentation/WordPerfect_Office_2000/Feature_Guides/fg-wp-2k-win-stand-int.pdf)

WordPerfect 9, which is an integral of WordPerfect office 2000, is an XML editor with extensive functions. No evaluation version was available on the Corel web site. A White Paper published in 1999 describes the available functionality.

4. EMACS add-on for SGML (http://www.lysator.liu.se/projects/about_psgml.html)

The text editor is available for UNIX environments and a clone has been created for Windows NT. There is an add-on available to make SGML editing possible. The programme is freeware but requires significant technical expertise, both in installation and use, so it was not included in the evaluation.

5. Xeena (http://www.alphaworks.ibm.com/tech/xeena)

The editor takes as input a given DTD, and automatically builds a palette containing the elements defined in the DTD. Users can thus create/edit/expand any document derived from that DTD, by using a visual tree-directed paradigm. This is a Java product that for unknown reasons did not work in the test environment.

Short List Evaluation

The evaluation criteria were used to compare the short listed tools. These were divided into three types: formal criteria, editor functions and extra features. The following editors were compared.

Evaluated Tools in the Short List							
Name	XED	XMetal	XML Pro	XML Notepad	Architag	XML Spy	XML Writer
Company	University of Edinburgh (HCRC)	SoftQuad	Vervet Logic	Microsoft		ICON Information systems	Wattle
Version	0.5.3.2	2.0.3.068	2.01	1.5 Beta	?	3	1.21
Release date	10/3/2000	14/9/2000	19/9/2000	3/5/1999	3/3/1999	18/5/2000	1/5/2000
Trialware (days)	Evaluation only	30	No save and print	Yes	n.a.	30	30
Name	XED	XMetal	XML Pro	XML Notepad	Architag	XML Spy	XML Writer
Cost (USD)	n.a.	400	150	?	0	149	49
Platform	Win32, solaris, FreeBSD, Linux	Win 95/Win 98/NT	95/98/NTSolaris, Linux	Win 95/Win 98/NT	Win 95/Win 98/NT	Win 95/Win 98/NT	Win 95/Win 98/NT
Download size (Mb)	1,6	19	3	0,35	0,1	6	1,6
Requires MSXML parser	No	ForXLST	No	Optional, only on loading	Yes, IE5	No	Yes
Other require-ments	Python Included		Java R.E. (12 MB)				

The tools were compared on the ability to firstly edit an existing XML document and secondly to start a new document. The starting of a document on the basis of a Document Type Description (DTD) requires that it can be loaded or assigned to the new document. The editing can then either be done directly at the level of the source code or in a more structured environment. In general it is easier to edit in a structured environment but this needs not be the case if there is good context sensitive support by the software. All editors support the making of well-formed XML documents, but not all support validation. When getting to know an XML editor it is good if help and examples or sample files are available.

XML Editor Functions							
Name	**Loader DTD**	**Source code editing**	**Structured XML editing**	**Easy document editing**	**Validation supported**	**Help available**	**Examples**
XED	Yes	Yes	No	Yes	No	Yes	No
XMetaL	Yes	Yes	Yes	Yes	Yes	Yes	Yes
XML Pro	Yes	No	Yes	No	Yes	Yes	Yes
XML Notepad	No	No	Yes	No	(only on loading)	Yes	No
Architag	Yes	Yes	No	No	Yes	No	No
XML Spy	Yes	Yes	Yes	Yes	Yes	Yes	Yes
XMLWriter	Yes	Yes	No	No	Yes	Yes	Yes

XMetal and XML Spy support all the listed functions of the XML editor and prove to be the easiest software packages for editing XML documents. XED does not offer a structured XML editing or validation, but for a source code editor it is an effective tool which should be used together with a parser or Architag, which is an interface to the Internet Explorer parser.

The diagramme at the beginning of the chapter shows how XML editors can relate to other modules. The editors were evaluated to the extent that these links have been included. The first link is to a DTD editor, which was possible with some editors and actually integrated in XML Spy. Importing data for data conversion was possible with some products. Validating with a parser was either integrated or done with a link to the Internet Explorer parser. All products that supported previewing used Internet Explorer, although this was sometimes presented as a window within the editor.

Extra Features of XML Editors					
Name	**DTD editor included**	**Data conversion**	**Parser**	**Preview using a browser**	**Comments**
XED	No	No	No	No	
XMetaL	Optional link to XML Authority (50 US$)	Yes from ODBC with customisation possibilities	Yes	Yes	Customisation is possible for further user-friendliness, such as limiting functionality
XML Pro	Optional link to Microstar's near and far (150 US$)	No	Yes	No	The unregistered evaluation version does not save files
XML Notepad	No	No	No	No	User friendly editing but limited functions
Architag	No	No	Uses MSXML Parser (IE5.01)	No	There are no functions to save and load files, therefore it should be used in combination with a text editor
XML Spy	Yes	Yes	Yes	Yes	There is a table view for start of new XML documents
XMLWriter	Checks wellformedness	No	Uses MSXML Parser	Yes	Not always stable during testing

XML Spy was found to be one of the most useful XML editors. For general editing the features are similar to Xmetal, but the extra features that can be of use in further phases are more extensive. Besides editing DTDs and Schemas the editing of style sheets is also supported, even though this was not tested.

XML Spy does not support templates for entering the data, but blank files with a structure based on the DTD were created to contain the data. This in fact then works as a kind of template. Sections of the "blank files" which were not needed could easily be removed. This was easier than starting every file from scratch.

The structured view made it easy to browse through the data comparing and editing similar fields of different activities. This is not possible when viewing the source code as the data is presented sequentially according to the DTD.

The documents are only validated on request by the user or when loading and saving the file. The error messages given when a file was not valid in general helped find and correct problems effectively.

Editing XML in Practice

For a user who is interested in getting to know XML and the XML editors that are available there are a few criteria especially important when selecting an editor. These are:

- Platform availability
- Download size
- Sample files
- Cost

The editors do not support all operating systems so the user needs to select editors for the platform that is available. The editors are all available on Internet sites but differ significantly in size. When considering the download size one should also check what other programmes the editor requires. If Internet Explorer or a Java Run-time Environment is required and not yet available these are also available on the Internet but are quite big downloads.

To test the editors it is useful to have example XML files which are sometimes included with the downloads of editors. The sample files can of course also be used with editors that do not include sample files. Sample files can also be gained from other XML projects on the Internet.

When evaluating and getting to know XML editors the cost is not very important as all the editors evaluated in the short list had free evaluation versions available. After the trial period (usually 30 days) some of the editors may stop functioning or have limited functions. After the evaluation period one can consider buying an editor or using the free editors.

A Recommendation on XML Editors

If a user decides to work with XML editors, it is strongly recommended to look at more than one editor as justified above. A good start at present is XML Spy, which in this project was found to be one of the most useful XML editors. For general editing the features are similar to XMetal but the extra features, which can be of use in further phases, are more extensive. Besides editing DTDs and Schemas, the editing of style sheets is also supported even though this was not tested.

CONCLUSIONS

The conclusions are subdivided into four categories. Firstly on XML as a phenomenon, then on XML as a technology, thirdly on tools and finally on IDML as a specific international development related occurrence of XML.

XML: A New Phenomenon

Although very technology driven, a lot is happening around this theme and it has the new and challenging flavour of the Internet and new economy, both in terms of standardization and development of applications. The former can act as a strong catalyser in the up building of new forms of cooperation and information definition processes and the latter means that there are upcoming markets with strong business opportunities.

As XML concentrates on the semantics of text, it draws back the focus of the Internet community to the content, the meaning of text. This instead of a former line of development which was much more oriented towards presentation. As developing countries are often faced with a lack of bandwidth, and in general focused content is less bandwidth consuming than presentation, this is a very welcome development.

Strongly related to the first conclusion is the statement that XML is still a turbulent, unfinished development. This means that some of the lines of development that are currently initiated will never result in a mature product and as a result starting a XML business is a high risk business. Also it means that the current products (standards, applications) are unfinished, so that the actual usefulness can be low at present.

XML: A New Technology

XML is a powerful technology. Unicode, store and forward systems and mapping information to various devices are, especially for developing countries, some of the most interesting features or possibilities of this technology.

XML as a technology enables the sharing of information between organisations where DTDs can be taken from libraries of shared data definitions and XML advocates platform and application independence.

Technologically spoken, XML facilitates the autonomy of organisations that share information. This because a stakeholder can save and maintain its information on its own server, choose what to make public so that this can be spidered or harvested by an XML search engine. On the other hand however this requires decisions in the organisational sphere: it demands an active decision not to gather all data to the central level, as was more common in previous information systems, but instead follow this decentral model.

Working with XML: Tools

A process has been proposed in this document and tested to include a large number of possibilities but limit the testing required. This forms a sound basis for the repeating of this exercise at a later stage or when looking for products with a different functionality.

For a user who is interested in getting to know XML and the XML editors that are available especially important criteria are platform availability, download size, sample files and cost.

If a user decides to work with XML editors, it is strongly recommended to look at more than one editor as justified above. At present, a good starting point is XML Spy.

IDML: XML for the International Development Sector

Since IDML is XML, it shares the benefits of XML: it brings structure to information. On the other hand, the IDML DTD is only a limited set of tags, representing some, but not all relevant data on international development projects and therefore restricts what you can express.

The international development sector is a very information intensive sector with a lot of inter-organisational information exchange. This means that application of XML, as a means of exchanging structured information, in this sector is a logical step. This can go much further than the type of information at present possible within IDML.

By expressing information on different projects using the same (IDML) DTD, it becomes easier to compare these projects, between and within organisations.

IDML can be seen as a best practise for project information definition and thereby is also a useful starting point for the design of a data model for any (also non- XML) project information system.

The IDML documentation is at present incomplete. In fact there is almost no information on the DTDs. Most of the descriptions have to be derived from the older CEFDA documentation.

Up till now the IDML initiative is obviously ‘donor-driven’. As a result the current DTD mainly contains information that is relevant for donor organisations. Details on pipeline projects for instance are lost when expressed in IDML.

As mentioned above in the conclusions on technology, XML and thus IDML facilitate autonomy. This is especially relevant in the international development sector with so many different stakeholders, each having its own interest and involvement. Current initiatives on IDML however do not use this possibility but collect all the IDML data and store it centrally from where it is available for searching and publishing.

APPENDIX A: LONG AND SHORT LISTS OF TOOLS

This appendix contains the intermediate products during the process of the tool comparison. The information was collected from four different sources and the quality depends on these sources, which may be somewhat outdated. The first two tables include the available descriptions and the Internet resources of the short listed editors. The final table is a list of the editors that were not included in the short list.

The sources of the editors in the long list are the following four lists:

1. Investing In Knowledge Toolkit -- the results from an IICD sponsored research into XML for developing countries. This was done in the first half of 2000.

2. The Data Documentation Initiative -- a project for the documentation in XML of metadata. The format is being considered by a department of the World Bank which gathers data on household surveys done in Africa. The software list is part of the final report which includes evaluation of software for marking up data to XML/SGML.

3. Academic XML users in the Netherlands -- every year are a number of meetings parallel to an online discussion. One of these meetings (April 2000) was a demonstration of XML editors.

4. The Open Directory Project has a list of XML resources and specifically XML editors.

Name	Comment	List
Adept -	SGML/XML publishing tool for business and technical documents with word processor interface, support for multiple editors, and an API to allow interfacing with external repositories.	4
Adept Editor 8.0 (ArborText)	the editor costs $1500 and another piece of software called Document Architect ($7000) is also required for initial DTD loading.	2
Clip	?	1
CLIP -	Java-based validating editor with support for multiple authors, searching, and DTD structure viewing.	4
Clip! (Techno 2000, Alpha)	Most of these products have significant limitations.	2
Emacs + PSGML	Hans van Mourik (Univ Leiden - ETCL)	3
EMACS add-on for SGML	This programme requires a lot of specialized technical expertise	2
WordPerfect	?	1
WordPerfect Corel 2000	Caroline aan de Wiel (Univ Leiden - Let, VTW)	3
XED	no comment	2
XED	Rik Hoekstra (Inst Ned Gesch)	3
XED -	Text-based non-validating XML editor.	4
Xeena	Rik Hoekstra (Inst Ned Gesch)	3
Xeena -	Xeena is a generic Java application for editing valid XML documents derived from any valid DTD. (IBM Corporation)	1
Xmetal	excellent but expensive	2
XmetaL -	A document-centric authoring tool for XML and SGML.	4
XmetaL (SoftQuad).	This is the programme we ultimately used for markup. We were able to load the DTD with a minimum of effort	2
XMetal SoftQuad	Hans van Mourik (Univ Leiden - ETCL)	3
XML <PRO> (Vervet Logic)	Most of these products have significant limitations.	2
XML Pro -	A Java based XML editor that supports DTD validation.	4
XMLpro	no comment	1
XML Notepad	good	1
XML Notepad -	This utility shows how a simple editor can be built based upon the MS XML parser in IE4 and IE5.	4
XML Spy 3.0 - XML Editor -	XML Spy is an Integrated Development Environment (IDE) for XML that includes major aspects of XML in one product: a validating XML editor, a Schema editor with schema validation, and an XSL editor.	4
XMLSPY	my favourite	1
XMLSpy	Dagmar Stiebral (NIWI)	3
XMLSPY	Most of these products have significant limitations.	2
XML Toybox -	How to use MSIE 5 as a validating XML editor.	4
XML Toybox Architag	good for freeware	1
XMLWriter	Peter Boot ()	3
XMLwriter -	A validating XML text editor for Windows that supports XML, XSL, DTD, CSS, and HTML files.	4

Short List of Internet Links

The quality of these links is sometimes dubious, as they have been taken directly from the lists. For the up to date links see the documentation on the short listed tools in the chapter where they are compared.

Name	Internet Resource	List
Adept -	http://www.arbortext.com/editor.html	2
Adept Editor 8.0 (ArborText)	http://www.arbortext.com/Products/products.html	1
Clip	http://www.t2000-usa.com/	4
CLIP -	http://www.t2000-usa.com/product/clip_index.html	2
EMACS add-on for SGML	http://www.lysator.liu.se/projects/about_psgml.html	1
WordPerfect	http://www.corel.com/office2000/feature_guide.htm	4
XED	http://www.ltg.ed.ac.uk/~ht/xed.html	4
XED -	http://www.ltg.ed.ac.uk/~ht/xed.html	2
Xeena -	http://www.alphaworks.ibm.com/tech/xeena	2
Xmetal	http://www.softquad.com/products/xmetal/xm1-intro.html	4
XmetaL -	http://www.xmetal.com/	2
XmetaL (SoftQuad).	http://www.sq.com/products/xmetal/	1
XML Pro -	http://www.vervet.com	2
XMLpro	http://www.vervet.com	4
XML Notepad	http://msdn.microsoft.com/xml/default.asp	4
XML Notepad -	http://msdn.microsoft.com/xml/notepad/intro.asp	2
XML Spy 3.0	http://www.xmlspy.com/xml_editor.html	2
XMLSPY	http://www.icon-is.com/prev/index_prev.asp	4
XML Toybox -	http://architag.com/xmlu/play	2
XML Toybox Architag	http://www.architag.com/xmlu/play/editor/	4
XMLwriter -	http://XMLwriter.net/	2

Non-Short Listed Editors

These are the editors on the long list that did not pass to the short list. Some of these may actually not refer to editors but where categorised as such in the original lists. These editors or resources were not checked during the process.

List	Tool	Platform	URL	Comment from list
2	Balise		http://www.balise.com/	We have not yet had a chance to evaluate automated markup tools like Omnimark and Balise. These tools require a fair amount of technical skill to use. for converting legacy documents into SGML/XML structured form. Balise is the French word for markup
1	DXP	UNIX	http://xdev.datachannel.com/	?
4	Editml -		http://www.editml.com	A Windows hosted, validating, editor for data files, schema files and stylesheets. It is based on the Microsoft MSXML parser and claims conformance to the W3C XML specification 1.0.
4	EditTime -		http://www.timelux.lu/EditTIME3/EditTIME.htm	SGML/XML authoring tool with a focus on international applications.
4	Emil, -		http://www.in-progress.com/emile/	Emil, is optimized for PowerPC and includes a document outline palette, markup validator, contextual menus, context-sensitive elements lists and many other features to faciliate efficient markup.
4	epcEdit SGML / XML editor software -		http://www.tksgml.de	epcEdit is a full-featured (validating), multi-platform SGML/XML editor application based on a powerful SGML/XML library (called TkSGML).
4	EXML Editor -		http://www.cuesoft.com/products/exml.asp	Free non-validating editor for XML documents with simple tree- and source-views.
2	FrameMaker+ SGML 5.5 (Adobe)			We were unsuccessful in loading the DTD into the software. Adobe is promising that the next version of the software will be more user-friendly with more of an emphasis on XML. It is fairly expensive, $348
1	HEX	UNIX	http://www-uk.hpl.hp.com/people/ak/java/hex.html	?
1	Homesite	Win95/98/NT/UNIX	http://www1.allaire.com	Good
1	JUMBO	UNIX	http://ala.vsms.nottingham.ac.uk/vsms/java/jumbo/index.html	
1	LARK	UNIX	http://www.textuality.com/Lark	?
1	Larval	UNIX	http://www.textuality.com/Lark	?
2	MetaStar Repository (Blue Angel Techno-logies, Inc.):		http://www.blueangeltech.com/	Microsoft Access based structured document entry and edit tool.

List	Tool	Platform	URL	Comment from list
4	Morphon -		http://www.morphon.com/	This validating XML editor uses CSS as the styling language, supports context sensitive styling and has an API for storage plugins.
2	Omnimark			Programming language for text processing and web application development. Useful for bulk markup of documents with similar structure.
2	TXE - The XML Editor -		http://www.geocities.com/shjejurkar/TXE/readme.html	A GUI XML editor written in Java using the DOM (Document Object Model) parser provided by Oracle.
1	WebWriter	Win 95/98/NT	http://www.stilo.com/	no comment
1	XML	Win 95/Win 98/NT	http://alphaworks.ibm.com/formula/xml	no comment
2	XML Instance -		http://www.xmlinstance.com	Create, edit, and manage data-oriented XML documents, messages, and configuration files.
2	XML Styler -		http://www.arbortext.com/xmlstyler/	a tool for creating and modifying XSL stylesheets.
1	XMLPROC	UNIX	http://www.stud.ifi.uio.no/~larsga/download/python/xml/xmlproc.html	no comment
4	XmlTool Multi-purpose Tool - http://xmltool.tripod.com			Mouse based shareware editor for XML files, that generates HTML, XSL, DTD files, and gets Database structure and queries.
4	XMLtree -		http://b62.tripod.com/vbwork/xmltree.htm	A visual parser for XML files which allows editing the document using a visual tree metaphor.
4	Zveno Swish -		http://www.zveno.com/zm.cgi/in-products/in-swish/	A non-validating XML document editor with Tcl/Tkl and DOM scripting that allows the user to view and edit a XML document in both a tree-mode and a document-mode simultaneously.

APPENDIX B: XML RESOURCE PAGE

Introduction

XML in 10 points (*http://www.w3.org/XML/1999/XML-in-10-points*)
XML in 10 points (7, really...) XML, XLink, Namespace, DTD, Schema, CSS, XHTML,... If you are new to XML, it may be hard to know where to begin. This summary in 10 points attempts to capture enough of the basic concepts for a beginner.

XML cover pages (*http://www.oasis-open.org/cover/siteIndex.html*)
The following expanded contents listing serves as a hierarchical subject map for the entire SGML/XML Web Page.This expanded outline is useful for string searches and for navigation if one is not already familiar with the structure of the site.

Spotlight on XML -CNET Builder.com
(http://www.builder.com/Authoring/XmlSpot/?tag=st.cn.sr1.ssr.bu_xml)
You may have heard that XML, the relatively new Extensible Markup Language, will revolutionize the Web, and that XML has the potential to reshape the way data is exchanged over the Internet. Before you abandon HTML, though, find out what XML really means

Investing in Knowledge Toolkit (*http://www.oneworld.org/thinktank/iktools/edit2.html*)
XML for developing countries. An editorial and tutorial

W3C the place for XML standards (http://www.w3.org/XML/)
The official site which does not always make for easy reading.

BUILDER.COM - Web Authoring (http://www.builder.com/Authoring/Xml20/ss05.html)
20 questions on XML: For example: 5. What is a Document Type Definition (DTD)?

Web Design - XML - WITS_END (http://www.wits-end.com/bookmarks/web-xml.htm)
The Extensible Markup Language (XML) is a data format for structured document interchange on the Web. (extensible because it is not a fixed format like HTML). Both HTML and XML are subsets of SGML.

FAQ with answers (http://www.ucc.ie/xml/)
Frequently Asked Questions about the Extensible Markup Language. Last updated July 2000.

Examples

Examples of working implementations using XML:

The IICD/OneWorld Demonstrator (http://nt.oneworld.nl/xml/query.cfm)
This facility uses the Extensible Stylesheet Language to integrate data from the various sources. The only browser that supports XSL is Internet Explorer 5.0.

Association of Shareware Professionals -PAD (http://www.asp-shareware.org/pad/padgen.asp)
The PADgen tool produces Portable Application Data XML files to describe software. This can be used for different marketing purposes.

The Data Documentation Initiative (http://www.icpsr.umich.edu/DDI/codebook.html)
The DDI is an effort to establish an international criterion and methodology for the content, presentation, transport, and preservation of "metadata" (data about data) about datasets in the social and behavioral sciences.

IDML

International Development Markup Language.

IDMLinitiative (http://www.idmlinitiative.org/)
A participatory approach to improve information sharing, using XML, within the international development community.

IDML - Participants List (http://www.idmlinitiative.org/index.cfm?fuseaction=ShowTemp&template =/cont/dsp_organisations)
Participants List The IDML Initiative is an informal group of development professionals who are committed to improving access to information that is relevant for their work.

Integrated Development Activity Information (GDG) (http://gateway.arsdigita.com/idml/ui.tcl)
Integrated Development Activity Information (IDAI) of the Global development gateway

DTD proposals (http://www.gcw.nl/idml002)
The first DTD that Ron Davies has compiled in 1998 has never been officially endorsed.This DTD was very close to the CEFDA standard. In 2000 a number of draft DTD's were developed for data exchange in the GDG / IDAI pilot.

GDG Initial Pilot (http://www.idmlinitiative.org/gdg/)
The first pilot for the Global Development Gateway which uses GoXML as search engine.

Tools

XML Spy - Editor (http://www.xmlspy.com/schema_editor.html)
XML Spy is an Integrated Development Environment for the XML that includes all major aspects of XML in one powerful and easy-to-use product. It is centered around a professional validating XML editor that provides four advanced views on your documents.

XML Browsers at XMLSOFTWARE (http://www.xmlsoftware.com/browsers/)
XML Browsers Software for viewing XML documents. XML Browsers are generally driven by stylesheets, are tree-based or are application specific.

XED: An XML document instance editor (http://www.ltg.ed.ac.uk/~ht/xed.html)
A free evaluation text editor for XML document instances. It is designed to support hand authoring of small-to-medium size XML documents, and is optimised for keyboard input. It works very hard to ensure that you cannot produce a non-well-formed document.

Opera Browser (http://www.opera.com/)
Opera Software - Bringing speed and fun to Internet browsing. Already supporting some XML.

Google: XML Tools
(http://directory.google.com/Top/Computers/Data_Formats/Markup_Languages/XML/Tools/)
The Open Directory Project XML Tool Pages as offered by Google. Google Web Directory - Computers > Data Formats > Markup Languages > XML > Tools

Extensibility - the makers of XML authority (http://www.extensibility.com/)
Extensibility's XML schema design and conversion solution, XML Authority leads the industry. The all-new XML Authority 2.0 is now available!

Architag - a basic XML editor (http://architag.com/)
Architag uses Internet Explorer 5 to check if XML documents are created correctly

Dicussions

Bellanet discussion lists (http://www.bellanet.org/lyris/helper/index.cfm?fuseaction=Home)
Email Forums Hosted by Bellanet. Archives of public lists are available including idml-l International Development Markup Language Initiative.

XML users in the Netherlands (in Dutch)
(http://www.lsoft.com/scripts/wl.exe?SL1=XML-NL&H=NIC.SURFNET.NL)
Mostly academic users exchange experiences.

XML.com: Mailing Lists (http://www.xml.com/pub/rg/3)
Various XML and related technology discussion lists.

Related Technologies

Resource Description Framework (RDF) (http://www.ilrt.bris.ac.uk/discovery/rdf/resources/)
XHTML Basic (http://www.w3.org/TR/xhtml-basic/)
Xlink (http://www.w3.org/TR/xlink/)
A Resource Guide (http://www.xml.com/resourceguide/)
Extensible Stylesheet Language (XSL) (http://www.w3.org/TR/xsl/)
XSL is a language for expressing stylesheets. It consists of two parts: a language for transforming XML documents, and an XML vocabulary for specifying formatting semantics.

GLOSSARY

CEFDA	Common Exchange Formats for Development Activity Information
DML	Development Markup Language, another name for IDML because this could be confused with the Islamic Document Markup Language
DTD	document type declaration
Dynamic XML	XML data generated on the fly by software/scripts from data in other formats
IICD	International Institute for Communication and Development
ICT	Information and Communication Technologies
IDML	International Development Markup Language (or Islamic Document Markup Language www.idml.org)
INDIX	International Network for Development Information Exchange
ITM	ITM management and informatics
M&E	Monitoring and Evaluation
XML	eXtensible Markup Language

REFERENCES

Association of Shareware Professionals. What is PAD?
http://www.asp-shareware.org

Bos, B. XML in 10 points
http://www.w3.org/XML/1999/XML-in-10-points

Data Documentation Initiative. Final Report
http://www.icpsr.umich.edu/DDI/DDIFinalReport.doc

Mourik, H van. Contribution to the XML-NL list about demonstrations on the 12th of April 2000
http://listserv.surfnet.nl/scripts/wa.exe?A1=ind0009&L=xml-nl

Munyati, M. Investing in Knowledge: XML Editorial and Tutorial
http://www.oneworld.org/thinktank/iktools/edit2.html

Schloss, B. 10 best bets for XML applications
http://www-4.ibm.com/software/developer/library/tenxmlapps/index.html

IICD PROFILE

The International Institute for Communication and Development (IICD) assists developing countries to realise sustainable development by harnessing the potential of information and communication technologies (ICTs). The driving force behind IICD activities is that local 'change agents' themselves identify and develop proposals for realistic ICT applications - local ownership forms the essential basis for sustainable socio-economic development.

Acting as a catalyst, IICD's three-pronged strategy is mainly delivered through a series of integrated Country Programmes.

First, IICD facilitates ICT Roundtable Processes in selected developing countries, where local stakeholders identify and formulate ICT-supported policies and projects based on local needs.

Second, working with training partners in each country, Capacity Development activities are organised to develop the skills and other capacities identified by the local partners.

Third, IICD draws on its global network to provide information and advice to its local partners, also fostering local information exchange networks on the use of ICTs for development. The best practices and lessons learned are documented and disseminated internationally through a Knowledge Sharing programme.

In support of these activities, IICD invests in the development of concrete partnerships with public, private and non profit organisations, thus mobilising knowledge and resources needed by IICD and its local partners.

Country Programmes are currently being implemented in Bolivia, Burkina Faso, Ghana, Jamaica, Mali, Tanzania, Uganda and Zambia.