

Standardising Document File Format

Stephen McGibbon, Senior Director, EMEA Technology Office, Chief Technology Officer, Eastern Europe, Russia & CIS, Microsoft, UK

In late November a small news item caught my eye. It reported that Gemplus International had been awarded the Sesame Award for Best Software at the Cartes 2005 Conference and Exhibition for their newly released smart.NET card¹. My attention was drawn to this because Gemplus' smart .NET-based platform provides a compact but compliant implementation of the ISO/ECMA-335 standard – the international open standard which characterizes the .NET environment.

Now, when I explain to people that the .Net CLI (Common Language Infrastructure) and C# programming language are international open standards, surprise is by far the most typical reaction. So, for those who reading this may also be surprised (and with apologies to those already fully aware of the fact), here is a brief history²:-

- In August, 2000, Microsoft, Hewlett-Packard and Intel co-sponsored the submission of specifications for the Common Language Infrastructure (CLI) and C# programming language to the international standardization organization ECMA.

- ECMA formed two task groups (TG3 and TG2, respectively) within TC39, its technical committee responsible for programming languages and application development.

- The following organizations participated in the work of ECMA TC39/TG2 and TC39/TG3: Borland, Fujitsu Software Corporation, Hewlett-Packard, Intel Corporation, IBM Corporation, ISE, IT University of Copenhagen, Jagersoft, Microsoft Corporation, Monash University, Netscape, Novell/Ximian, Phone. Com, OpenWave, Plum Hall, Sun Microsystems, and University of Canterbury (NZ).

- During the next year, the co-sponsor companies, in conjunction with other ECMA members and guests (including IBM, Fujitsu Software, Plum Hall, Monash University and ISE), refined these specifications into standards.

- In December, 2001, the ECMA General Assembly ratified the 1st edition of the C# and CLI standards as EC-

MA-334 and ECMA-335, respectively.

- In late December, 2001, ECMA submitted the standards and TR to ISO/IEC JTC 1 via the latter's Fast-Track process.

- In April, 2003, ISO ratified the standards as ISO/IEC 23270 (C#), ISO/IEC 23271 (CLI) and ISO/IEC 23272 (CLI TR).

- In June 2005 the Ecma General Assembly approved³ the third edition of the standards for C# and CLI.

The news story of Gemplus' success stood out to me because I had recently spent a day talking to analysts and journalists about Microsoft and partners⁴ announcement⁵ to follow the same Ecma International/ISO path with the Office OpenXML formats for Word, Excel and PowerPoint. Gemplus and others success using the .Net CLI and C# standards is a good harbinger of Microsoft's intentions with the standardisation of the office file formats – we want them to be used.

The December 8th 2005 Ecma International General Assembly held in Nice created⁶ technical committee 45 (TC45) to produce a formal standard for office productivity applications that is fully compatible with the Office OpenXML formats, submitted by Microsoft with the aim of having this standard implemented on a wide set of tools and platforms.

The press release reads:-

Opening Up Billions of Documents – By standardizing and fully documenting the Office OpenXML formats, billions of existing Office documents can be converted to the Ecma XML standard format with no loss of data, for-

matting, properties, or capabilities. This will facilitate document contents access, search, use, integration and development in new and innovative ways. Documents and public records can also be created, archived, and maintained with long-term, widespread industry support.

“Documents are a fundamental means of personal and business communication, and the foundation of information archival,” said Jan van den Beld, Secretary General of Ecma International. “The benefit of backward compatibility that this effort will provide is evident from the broad spectrum of sponsors that will work together under the formal Ecma standardisation process. Following publication in Ecma, the standard is to be submitted for adoption under the ISO/IEC JTC 1 process.”

The substantial initial effort to include specifications for existing document types will amount to several thousands of pages. Additionally, TC45 will be responsible and be the governing body for enhancing the standard with new and innovative formats and features while maintaining backward compatibility. The third main responsibility of TC45 will be the ongoing maintenance of the standard. Participants are hopeful that they will complete the first edition of the standard before the end of 2006.

Microsoft expects that a new industry sector will grow around the standards that emerge from TC45, with representation from the whole software development spectrum from “free” through hybrid and open source through purely commercial. Indeed, to encourage exactly this broad adoption and support, Microsoft has taken a licensing approach that goes substantially beyond any Open Standards norms or requirements and created an “Irrevocable covenant not to sue” (which is posted at <http://www.microsoft.com/office/xml/covenant.mspx>). Whilst you should use that link for the official statement, I've also copied the text below:-

Microsoft Covenant Regarding Office 2003 XML Reference Schemas

Microsoft irrevocably covenants that it will not seek to enforce any of its patent claims necessary to conform to

the technical specifications for the Microsoft Office 2003 XML Reference Schemas⁷ posted at <http://msdn.microsoft.com/office/understanding/xmloffice/default.aspx> (the “Specifications”) against those conforming parts of software products. This covenant shall not apply with respect to any person or entity that asserts, threatens or seeks at any time to enforce a patent right or rights against Microsoft or any of its affiliates relating to any conforming implementation of the Specifications.

This statement is not an assurance either (i) that any of Microsoft’s issued patent claims cover a conforming implementation of the Specifications or are enforceable, or (ii) that such an implementation would not infringe patents or other intellectual property rights of any third party.

No other rights except those expressly stated in this covenant shall be deemed granted, waived or received by implication, or estoppel, or otherwise. In particular, no rights in the Microsoft Office product, including its features and capabilities, are hereby granted except as expressly set forth in the Specifications.

The question is, of course, whether this covenant will be accepted by the Open Source community. Happily, it appears that this will be the case. In a recent story entitled “Top open source lawyer blesses new terms on Microsoft’s XML file format”⁸, ZDNet’s David Berlind reported that Larry Rosen, “the attorney that wrote the book on open

source licensing and the man who was the Open Source Initiative’s first general counsel and secretary” had issued the following statement endorsing the new terms under which Microsoft is making its Office OpenXML formats available to developers of all types, including open source developers.

“I was delighted to learn of Microsoft’s recent “Covenant Regarding Office 2003 XML Reference Schemas.” This covenant goes beyond anything Microsoft has ever done before. It means that both open source and proprietary software can compete in implementations of these important XML schemas without the threat of patent litigation from Microsoft.

This covenant is at least as generous as the patent licenses for many other document formats and industry standards. It includes protection for Microsoft against patent lawsuits; this is just like the patent defense provisions in many open source licenses. And the scope of their patent covenant, even though it is limited to “conforming” software products, is sufficient to allow open source implementations that can read and write Office 2003 documents. Microsoft’s covenant is, to coin a phrase, as fair and balanced as other licenses or covenants we’ve accepted before. I am pleased to see Microsoft move their patent licensing strategy this far.

Microsoft has offered its specification for standardization by ECMA, an industry standards organization headquartered in Europe. It is important for

open source companies to participate in this standardization effort, so that we can ensure that the specification for the standard is itself developed in an open way. If we do that, I’m confident that “conforming” software products will evolve to meet customer needs worldwide without Microsoft having to dictate the scope of that conformance.

The first reaction people will have is, “where’s the catch?” I don’t see anything we can’t live with. We can participate in crafting the standard in ECMA, we can read and write Office 2003 files in open source applications, and we don’t have to pay royalties to Microsoft to do so. It’s a good start.”

Berlind goes on to say that Rosen is one of two lawyers whose endorsement is critical to the open source world’s acceptance of the Office OpenXML formats and that Rosen’s endorsement “apparently means that the open source community is free to develop software that supports the Redmond, WA-based company’s XML-based file formats for its Office productivity suite.”

Rosen was right about many people in the Open Source world’s first reaction. Many thousands of words have been written questioning Microsoft’s motives and building conspiracy theories about what our real agenda is. Responding to some of these concerns Rosen further commented⁹:

“Free software and open source software can forever implement certain Ecma-standard and OASIS-standard XML document formats. That’s what I insisted upon in every conversation I had with Microsoft’s attorneys and Sun’s attorneys during all our past quarrels over patent license terms. I believe that’s what we got from both companies.

That’s why I’m pleased, even though I’m not naive enough to believe there aren’t subsequent battles we’ll have to fight about open standards. The best way to ensure ongoing cooperation is for open source companies to join the standardization efforts at Ecma and OASIS. Keep them open.”

As I have laid out above, there’s also the positive history of C# and the .Net CLI to look to. Novell certainly seem to think the process is working and that the standard is real and useful. David Patrick, Novell’s general manager of

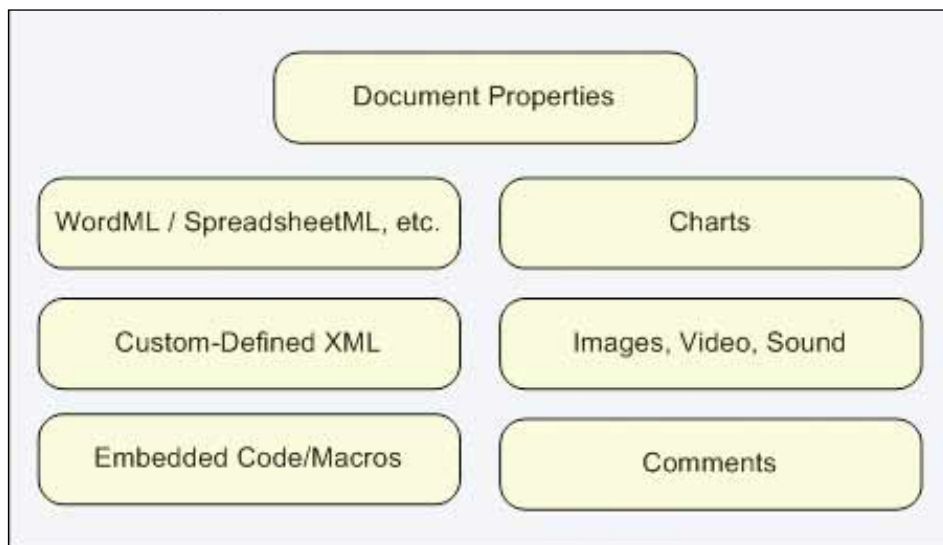


Figure 1. OpenXML format file container

Linux, open source and platform services is quoted in the June 2005 Ecma press release as saying

“We are very pleased with the upgrades to the Ecma C# language specification as they show the evolution of a language which is adapting to the needs of modern applications.”

So, with Office Open XML document formats available as an open standard, customers will have even more confidence in their ability to store and manage data for the long term, with many more vendors and tools from which they can choose. The move will benefit the broader software ecosystem because software and services vendors worldwide will be able to more easily build compelling solutions that interoperate across a broad spectrum of technologies.

This was really brought home to me by Stephen Deakin, director for information technology at Barclays Capital, one of the co-sponsoring partners who participated in the press and analyst briefing calls. Deakin explained that Barclays are particularly interested in the Open XML Format for spreadsheets, and used “live” bank statements with pivot tables as an example. These could be created by Barclays back office servers without the need for Microsoft Office.

Another representative of the co-sponsoring partners was Adam Farquhar, head of e-Architecture for the British Library who noted that the BL expected that “establishing Microsoft Office Open XML as an open standard will substantially enhance our ability to preserve and provide access to the U.K.’s digital heritage. It’s an important step forward for digital preservation and will help us fulfil the British Library’s core responsibility of making our digital collections accessible for generations to come.” Of course National Libraries across the world share a similar objective and this standardisation is equally important and beneficial to all. In the past scientists and artists would bequeath a set of notebooks to their national archive, nowadays they are far more likely to leave digital documents be they spreadsheets from experiments, drafts of manuscripts, or other media files.

As its name suggests the Office OpenXML format is based on the popu-

lar XML standard for data exchange making document-based data more accessible amongst heterogeneous systems. The openness of the Open XML Format unlocks data and introduces a broad, new level of integration beyond the desktop. Server-side applications will be able to process documents in bulk to enable large-scale solutions that mesh enterprise data within the familiar, flexible Office applications. Standard XML protocols, such as XPath (a common XML query language) and XSLT (Extensible Stylesheet Language Transformations), will be able to retrieve data from documents or to update the contents inside of a document from external data.

Importantly, the Office OpenXML format supports the use of custom XML schemas. This gives developers and organisations an important tool to share and re-use their own data. Custom schemas not only make insertion or extraction of data simple, but they also add structure to documents and are capable of enforcing data validation. The European Commission’s IDA committee recognised the advantages of custom schema support in Microsoft’s earlier XML schemas and asked Microsoft to consider submitting this work for standardisation. At that time the specifications were not ready to be submitted as they did not allow for the complete representation of the binary formats – something that ironically led to accusations Microsoft was being less than open. In fact Microsoft had been on this path for several years, and the schemas submitted to Ecma will represent all of the features in Word, Excel and PowerPoint.

The decision to submit the Office OpenXML formats for standardisation and the subsequent creation within Ecma International of TC45 represent an important watershed in the history of personal computing. Already the de-facto standard, Office formats will be available to the full spectrum of developers in a new full fidelity XML representation benefiting interoperability as there is less ambiguity and reduced scope for differing interpretation by developers.

The Office OpenXML format will be the default format in the forthcoming Office 12 release, and Microsoft has

committed to release upgrades to allow Office 2000, XP and 2003 to also be able to read and write the OpenXML formats. Microsoft will also make tools available to bulk convert binary formats to the OpenXML formats.

As Jean-Philippe Courtois, president of Microsoft International said at the announcement, “We are committed to open standards such as XML to provide the highest levels of interoperability between legacy and next-generation software. The creation of an XML file format standard is a major industry milestone. We hope this will provide both users and organizations with the peace of mind that they will be able to access their past and future documents for generations to come.” □

¹ http://www.gemplus.com/press/archives/2005/id_security/15-11-2005-NET_sesame.html

² <http://msdn.microsoft.com/netframework/ecma/>

³ http://www.ecma-international.org/news/PressReleases/PR_TC39_CSharp_CLI.htm

⁴ The Ecma members sponsoring the submission are Apple, Barclays Capital, BP, the British Library, Essilor, Intel Corporation, Microsoft, NextPage Inc., Statoil ASA and Toshiba. Other Ecma International members will be able to participate in the working group.

⁵ <http://www.microsoft.com/presspass/press/2005/nov05/11-21EcmaPR.mspx>

⁶ http://www.ecma-international.org/news/PressReleases/PR_TC45_Dez2005.htm

⁷ Notice that the text currently says that it applies to the Office 2003 schemas because we haven’t submitted the Office 12 schemas yet. The language will change to reflect that submission.

⁸ <http://blogs.zdnet.com/BTL/?p=2192>

⁹ <http://www.groklaw.net/article.php?story=20051202135844482>