

# **The Future is Open:**

*What OpenDocument is and why you should care*

*Daniel Carrera.*

*OpenOffice.org volunteer.*

## Table of Contents

What is OpenDocument?.....	3
Executive summary.....	3
Time-line.....	3
The European Union and the Valoris report.....	3
Selection criteria.....	4
Neutrality.....	4
Technical merit.....	4
Widely adopted.....	4
Finalists.....	5
OpenDocument vs MS XML.....	5
Reactions to the report.....	5
Microsoft responds to the Valoris report.....	5
Sun responds to the Valoris report.....	5
The EU TAC makes recommendations.....	6
Microsoft responds to the TAC recommendations.....	6
Sun responds to the TAC recommendations.....	7
IBM responds to the TAC recommendations.....	7
Where we are, where we need to go.....	7
This is a battle we can win.....	7
What OpenDocument needs to succeed.....	8
What you can do to help.....	8
Acknowledgments.....	9
Resources.....	9
Appendix I – Expanded time-line.....	10

# The Future is Open:

## What OpenDocument is and why you should care

### What is OpenDocument?

#### Executive summary

**Quiz:** Complete the sentence, “OpenDocument is ...”

- (a) An open, XML-based file format.
- (b) An open standard, supported by the OASIS and ISO standards groups.
- (c) The default file format for the upcoming OpenOffice.org 2.0 and KOffice 1.4.
- (d) A top prospect for an official format for the European Commission.
- (e) Our best chance to fight vendor lock-in associated with proprietary formats.
- (f) All of the above.

The correct answer is (f) All of the above.

#### Time-line

This article links to reports written over a period of several months. To make things easier to follow (in particular, the name of the format), I include an approximate time-line.

- 1) OpenOffice.org creates an open file format called “**OpenOffice.org 1.0**” format.
- 2) The European Union commissions Valoris to report on open file formats.
- 3) The “OpenOffice.org 1.0” format is submitted for OASIS standardization. KDE and Corel join the OASIS Technical Committee and expand the format to cover a wider range of applications.
- 4) The new OASIS format is called “**Open Office XML**”. OpenOffice.org and KOffice both commit to making the format their primary/native format.
- 5) The Valoris report is published. Microsoft and Sun respond to it. The European Union TAC makes recommendations.
- 6) The format is submitted for ISO standardization and changes its name to the “**OpenDocument**” format.

In order to avoid promoting an outdated name, I will always refer to the format as “OpenDocument”.

### The European Union and the Valoris report

The key player here is the European Union's “Telematics between Administrations Committee” (TAC). The TAC hired the Valoris consulting group to explore the possibility of using Open Standards to create a more competitive marketplace.

The Valoris report was well researched and very thorough. They recommended the

adoption of an open XML format backed by a recognized open standards group.

The Valoris report is a very detailed and most compelling document describing how open XML technologies are impacting government information systems and requirements policies.

It's difficult to summarize a 78-page document. I encourage you to [read it yourself](#). But over all, it is well researched and rather encouraging of open source, OpenDocument, and EU hopes for an open marketplace.

The Valoris group closely monitored the development of the OpenDocument format, and by November 2004 every one of the concerns and issues they raised were directly addressed by OASIS TC. As a case in how governments and organizations go about setting information system requirements policy, the Valoris effort provides an unparalleled insight.

### **Selection criteria**

For the report, Valoris compiled an exhaustive list of existing file formats. These were compared against a series of requirements. Only two formats performed well, and these were selected for further analysis.

Being familiar with these criteria is important for us. It teaches us what a government requires before it will adopt an open XML format. I divide these into three groups:

#### ***Neutrality***

- ◆ Open: At this stage of analysis Valoris only looked for openness in the sense that a public, royalty-free specification is available. For example, PDF and MS XML both met this requirement.
- ◆ Non-binary: Binary formats get in the way of neutrality. A format that depends on Windows components will be hard to support on GNU/Linux and Mac OS.
- ◆ Cross-platform: An obvious requirement for neutrality.

#### ***Technical merit***

- ◆ Preserve format fidelity: This refers to both presentation and structure. For many applications, format fidelity is an absolute imperative.
- ◆ Modifiable: This excludes formats like PDF.
- ◆ Support current word processor features: An open format is no use if it can't represent your data. This list included Unicode support, bi-directional (Hebrew), and scripting, among others. Abiword and KOffice failed this requirement.
- ◆ Support emerging requirements: Digital signatures, access rights, version control, etc. Almost every format failed this requirement.

#### ***Widely adopted***

This sounds odd. Why is this important? Shouldn't we pick the best format regardless?

The answer is simple: Have you heard of the ODA? I didn't think so.

In the 80's Europe made an attempt to define an open standard format called *Open*

*Document Architecture* (ODA). It received a lot of support from European institutions including the European Commission. It was also an ECMA standard and an ISO standard.

ODA failed miserably.

ODA had no backing from the industry. The format was complex, and the companies preferred to support more pragmatic standards such as SGML and RTF.

Conclusion: Industry support matters. This is why the W3C never accepts a standard unless it has at least one working implementation.

*Note: “widely adopted” does not mean “dominant”. It means that there is enough industry adoption to sustain the format. For example, OpenDocument meets this requirement. Unlike the ODA effort, OpenDocument enjoys the support of two office suites as well as major players like Sun, IBM, HP, Novell and even Adobe.*

## **Finalists**

No format met all the criteria perfectly well. But two formats stood out above the rest:

- ◆ OpenDocument – main failing: lacked some support for emerging requirements.
- ◆ Microsoft XML – main failing: medium fidelity, unsure about cross-platform status.

## **OpenDocument vs MS XML**

The only apparent advantage of MS XML over the OpenDocument format was the presence of custom-designed schemas. On the other hand, OpenDocument had several advantages:

- ◆ More open: No legal constraints, and support from OASIS.
- ◆ Reuse of existing open standards when possible (SVG, Dublin Core, MathML, etc).
- ◆ Higher format fidelity.
- ◆ Friendly to XSLT and other XML-based tools.

## **Reactions to the report**

### **Microsoft responds to the Valoris report**

[Microsoft's response](#) essentially comes down to trying to convince the reader that custom designed schemas are really really important. This isn't surprising since it seems to be the only card they could play.

### **Sun responds to the Valoris report**

[Sun's response](#) expresses several concerns about Microsoft's approach to XML. I recommend reading the full response for details, but here are a few:

- ◆ MS XML doesn't cover all of MS Office (for example, it doesn't cover PowerPoint).
- ◆ MS XML doesn't support some of the advanced MS Office features.
- ◆ MS XML can contain binary objects that depend on MS Office and Windows (e.g. OLE and VBA) and **those** lack complete documentation.

- ◆ MS did not commit to make future changes to MS XML available to the public, only the current one.

## The EU TAC makes recommendations

After getting responses, the EU TAC wrote their [final recommendations](#). Those too are very encouraging. They stood firm on the importance of file formats:

*“...the public sector should avoid any format that does not safeguard equal opportunities to market actors to implement format-processing applications, especially where this might impose product selection on the side of citizens or businesses...”*

Their response to Microsoft's argument on custom-defined schemas was interesting. Instead of arguing whether the feature is really valuable, they chose to challenge OASIS to add custom-defined schemas to the format to match Microsoft's argument. The TAC also had some suggestions for Microsoft, finding that the MSXML and MS XML Reference License were simply not “open enough”.

Recommendations for OASIS:

- ◆ Add custom-defined schemas to the format.
- ◆ Submit the format to ISO for a more official seal of approval.

Recommendations for Microsoft:

- ◆ A public commitment to keeping future versions of MS XML public.
- ◆ Submit the format to an international standards body of their choice.
- ◆ Remove non-XML components from the format.

Recommendations for other industry players:

- ◆ Participate in the OpenDocument standardization process to encourage wider industry consensus around the format.
- ◆ Include filters to support both OASIS OpenDocument and MS XML.
- ◆ Provide tools to help the public sector migrate its documents to XML formats.

Recommendations for the general public:

- ◆ Provide your documents in multiple formats, or alternatively, in an open format with industry consensus and adoption.

## Microsoft responds to the TAC recommendations

[Microsoft's response](#) essentially comes down to:

- ◆ They agree to publish future versions of MS XML under “non-discriminatory” terms.
- ◆ They say they will “vigorously” work on documenting the non-XML elements of the format, though they disagree that they should exclude all non-XML elements. They claim that OpenDocument also has non-XML elements (such as images) but fail to point out that in OpenDocument those are in a separate directory and in MS XML they are embedded throughout the XML tags.

- ◆ They also say that supporting various formats is important, but fail to say that they'll support OpenDocument.
- ◆ They insist that their licensing scheme is just fine, and that “royalty-free licensing programs have a role to play alongside formal standards”. Translation: No, they won't submit it to a standards body.

### **Sun responds to the TAC recommendations**

[Sun's response](#) reiterates the importance of formats that are truly open, in every sense. They like the idea of ISO support and are confident that the format will become an ISO standard.

### **IBM responds to the TAC recommendations**

[IBM's response](#) was also positive. There was a tiny bit that concerned me:

*... we believe [that standard] technologies should be published without restriction (other than reasonable royalties for essential patents) ...*

Well, I don't really know much about this, so I'll let other people decide if the stuff in brackets is significant. But other than that, IBM informs the EU that:

- ◆ They will join the the OASIS Technical Committee.
- ◆ They already offer products (Workplace) that conform with OpenDocument.

This last bit isn't surprising, since Workplace is based on OpenOffice.org code, but it's a good reminder.

### **Where we are, where we need to go**

The OpenDocument format is arguably the single most important step for the FOSS movement after software patents. It is unique in that:

- ◆ It cuts right to the core of the fight: Vendor lock-in.
- ◆ This is a battle we *can* win.

### **This is a battle we can win**

The OpenDocument format is in a unique position because it has a real chance of succeeding:

- ◆ **Technical merit:** With the inclusion of custom schemas, OpenDocument can meet and exceed every technical hurdle that can be thrown at it.
- ◆ **Adoption:** OpenOffice.org, KOffice, IBM Workplace, StarOffice. OpenDocument is not a dream. It is a real format with enough support to present a real alternative.
- ◆ OpenOffice.org is cross platform, easy to download and install, and is preparing to ship version 2.0. Sun, IBM, HP, Red Hat, Novell, and even Adobe were able to announce full support of the EU open XML standards requirements based on the inclusion of OpenOffice.org components in their products and services. That both OpenOffice.org and KOffice are open source with open component frameworks enables the entire marketplace of vendors to easily comply if they choose to do so.

- ◆ **Open standard:** OpenDocument is more than an open format, it is an open *standard*. That is, it is backed by standards groups, ISO and OASIS. It is not controlled by any company or product. Not by Sun or IBM. Not by OpenOffice.org or KDE. Is Microsoft willing to match that?
- ◆ **Political will:** The European Union is pushing for a policy of open standards. So is the state of Massachusetts, USA. The time is right to push for OpenDocument as the standard for government documents. People are listening.

## What OpenDocument needs to succeed

The Valoris report makes it clear that any open format that is to succeed as an official standard for government documents needs four things:

- ◆ **Technical merit:** Most open formats that were evaluated did not make it to the final round because they lacked important features. Any technical shortcoming, no matter how theoretical, gives Microsoft a lobbying opportunity.
- ◆ **Adoption:** No adoption, no success. Learn from ODA.
- ◆ **Open standard:** Why is this so important? Why not just any open format like the old OpenOffice.org 1.0 format or KOffice? Because Microsoft would argue unfair treatment. Yes, I know, ha ha. But still, Microsoft could effectively lobby against any format controlled by one application supplier.
- ◆ This is why OpenOffice.org and KOffice are both switching to OpenDocument as their primary format (in versions 2.0 and 1.4 respectively).
- ◆ **Political will:** The fight for open standards will not be easy. It will take will power from politicians, *and the public who elects them*, to come through. Open XML technologies are the future of collaborative computing, but the proprietary efforts to compromise their implementations, or lock them up with insidious patents and restrictive licenses, continues.

## What you can do to help

Keep your focus on the four items above. Let's go through each and see what we need.

Technical merit

### Check

If you find a significant failing in the format, join the OASIS TC and participate. Or, you can follow the example of XML expert David Wheeler. Even though David is not a member of OASIS, he routinely participates in discussions and contributes to the specification through the public forum. The OASIS TC also maintains liaisons with other open XML efforts such as the OASIS UBL and the W3C XForms groups.

Then there is always the option of working through one of the many OpenOffice.org project groups to directly impact the specification. Examples of this include the Bibliography and DocBook groups who labored long and hard to prepare the specification to perfect the demands of the Library of Congress' very complex "MODS" schema.



Adoption	<ul style="list-style-type: none"> <li>• Write filters for Abiword, Gnumeric, Sodipodi and Inkscape.</li> <li>• Promote applications which support OpenDocument.</li> <li>• Help at OpenOffice.org or KOffice.</li> </ul>
Open standard	<b>Check</b>
Political will	<ul style="list-style-type: none"> <li>• Help convince the EU and MA to support OpenDocument.</li> <li>• Write letters, lobby.</li> <li>• Inform people. Feel free to use this article to do so.</li> <li>• Become informed yourself.</li> <li>• Read the Valoris report and the documents listed here.</li> <li>• If you live near San Diego, California, go to the OpenOffice.org RegiCon on Feb 9 and hear Gary Edwards from OASIS talk about OpenDocument:</li> <li>• <i>The Shot Heard Round the World: How OASIS Open Document Changes Everything.</i></li> <li>• Attend the above talk and provide a transcript for Groklaw.</li> </ul>

### **Acknowledgments**

Gary Edwards, from the OASIS TC, reviewed this article for technical accuracy. Jean Hollis Weber edited the final draft. Any errors or omissions are entirely my own.

### **Resources**

[Read the Valoris report, along with TAC recommendations and all responses.](#)

[OASIS Technical Committee.](#)

This article is licensed under the same [Creative Commons License](#) as PJ's own articles.

## Appendix I – Expanded time-line

This expanded time-line was provided by Gary Edwards.

- 1) Sun purchases the cross platform office productivity suite, "StarOffice". They create the open source community, "OpenOffice.org", and open source the entire code base of StarOffice.
- 2) OpenOffice.org creates an open XML file format that covers all compound document aspects of the productivity suite. It is the first time the volumes of desktop productivity information can be captured in a "structured" file format.
- 3) Sun and the OpenOffice.org community submit the open XML file format specification to OASIS as an open standards candidate.
- 4) The OASIS Open Office TC (Technical Committee) begins 18 months of work on phase one, which was completed in March of 2004. Phase one work is focused on meeting the needs of enterprise level content management and publication concerns. Contributors such as Boeing, Arbortext, Stellent, Documentum, Corel, SpeedLegal, the Australian National Archives, and the International Biblical Society lead the way in expanding the specification to meet the transformation demands of having over 30 years of legacy applications and systems still on line, and still productive. Phase one work is focused on scalability and inclusiveness. The file format specification is positioned as a common transformation layer between legacy information systems and the rapidly advancing publishing and content management systems of the future.
- 5) The European Union commissions Valoris to study two concerns. The first is that of information systems based on open standards. The second is that of how the EU can create an open and highly competitive marketplace of information technologies. Valoris quickly moves to focus on open XML technologies, and the events taking place at the OASIS Open Office XML file formats TC.
- 6) In Phase two, the OASIS TC focus shifts to the emerging needs of desktop productivity applications. KOffice joins the TC, and work begins in earnest on expanding the specification to include a wider range of open XML technologies and multi media formats. Initiatives for XForms, SVG, SMiL, Bibliography, UBL, XBR, DocBook, project and content management schemas, custom defined schemas, intelligent compound documents, workflow integration, and more are undertaken. The OASIS TC begins to think in terms of a "productivity environment" and considers the demands of a collaborative computing future.
- 7) The Valoris report is published. Microsoft and Sun respond to it. The European Union TAC makes recommendations.
- 8) The OASIS TC responds to the EU recommendations, including the custom-defined schema fundamentals. The format is submitted to OASIS for formal standardization, with ISO standardization to follow. In keeping with the neutral and open spirit of the EU requirements, and the rapid expansion of the specifications productivity environment capabilities, the TC changes name to the "**OpenDocument**" format.