

OPEN DOCUMENT FORMATS AS AN ENABLER OF INTEROPERABILITY
COMPARISON OF THE OASIS OPENDOCUMENT FORMAT
AND MICROSOFT OFFICE OPEN XML

PC productivity applications (PPAs), including word processing, spreadsheet, and presentation software, represent the most important applications to PC users today. Most documents today are created with a word processor, spreadsheet or a presentations application. Therefore, for the purposes of ensuring that the documents can be freely exchanged between users, and that they are accessible not only now but also in the future, it is important that the format the documents are saved in is one that can be fully supported by the widest possible range of applications. Only a fully open cross-platform format can allow for full interoperability between PPAs, and ensure that users' data is not locked-in to specific applications and platforms.

OpenDocument Format and Microsoft Office Open XML

The OpenDocument Format (ODF) was created to meet the demand for such a format. It was developed as an application-independent file format by a vendor-neutral standards developing organization OASIS, with the participation of multiple office application vendors. On 4 May 2006, the Joint Technical Committee for Information Technology Standards (JTC-1) of the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC) approved ODF as international standard ISO/IEC JTC 1 26300.¹ ODF is supported by various PPAs, including OpenOffice,² StarOffice 8³ Google Docs & Spreadsheets,⁴ IBM Productivity Tools⁵, KOffice,⁶ and Zoho Office.⁷ Corel Corporation has also announced that it will support ODF in its WordPerfect Office by the middle of 2007.⁸

¹ From ODF Alliance, "About the OpenDocument Format", available at
<http://www.odfalliance.org/resources/AboutODF.pdf>

² <http://www.openoffice.org>

³ <http://www.sun.com/software/star/staroffice/index.jsp>

⁴ <http://docs.google.com>

⁵ <http://www-142.ibm.com/software/sw-lotus/products/product4.nsf/wdocs/productivitytools>

⁶ <http://www.koffice.org>

⁷ <http://www.zoho.com>

⁸ <http://www.corel.com/servlet/Satellite/us/en/Content/1153321430604?pressId=1164741065876>

Microsoft's Office 2007 introduces a competing format, known as the Office Open XML format (OOXML). The new format is currently not supported by any other PPA suite except Microsoft Office. Novell and Corel Corporation have announced intended support for OOXML in products, with Corel aiming to support the OOXML by the middle of 2007 on its Windows-based WordPerfect suite in addition to its ODF support.⁹ OOXML was approved as an ECMA standard in December 2006, and is currently being considered for adoption as an ISO standard. Whereas ODF was designed to be a vendor-neutral format capable of being supported on multiple platforms, ECMA describes OOXML as being

"designed from the start to be capable of faithfully representing the pre-existing corpus of word-processing documents, presentations, and spreadsheets that are encoded in binary formats defined by Microsoft Corporation."¹⁰

However, even if an independent vendor application faithfully supports the OOXML specification, it will not be able to open files stored in Microsoft's binary formats – unless additional support for those binary formats is included. An opaque method of storing information in a file that is not readily human-readable, binary formats encode information as a series of 0s and 1s in a structured manner. Critically, the structure of Microsoft's binary formats is known only to Microsoft, and it has refused to disclose them to other PPA developers.

Comparison of ODF and OOXML

True interoperability is not limited to specific platforms. Thus the real threshold test for document interoperability is whether the document format in question can be fully implemented in any application, regardless of the operating system on which it runs

The merits of ODF have already been established by its wide industry adoption. As noted above, numerous PPA vendors have implemented support for it in their products both on Windows and on other operating systems. Such widespread adoption is only possible because ODF is fully disclosed and created to allow for document interoperability by making it easy for various applications to exchange documents with full fidelity, *i.e.*, without any loss of data or formatting of the document.

In contrast to the industry's ODF efforts, with OOXML Microsoft ostensibly steps towards openness, but in fact strikes another blow to the hopes of restoring competition and consumer choice in the operating system and PPA markets. Research by ECIS as well as the software community¹¹ has demonstrated that the OOXML file format cannot be supported by non-Microsoft applications with full fidelity on other operating systems such as Linux or Solaris. It seems that Microsoft's motives in partially opening up its PPA document formats are not so

⁹ In February 2007 Microsoft also announced support for ODF in Office through means of an ODF add-in converter for Word.

¹⁰ http://www.ecma-international.org/news/TC45_current_work/OpenXML%20White%20Paper.pdf

¹¹ For an excellent resource, see the EOOXML Objections Wiki, available at http://www.grokdoc.net/index.php/EOOXML_objections

much an attempt to ensure document interoperability in the industry, but to make it easier for Windows developers to implement support for the Office document formats in their applications, thus further strengthening the role of Office as one of the pillars of Microsoft's monopoly power.

In contrast to OOXML, the adoption of ODF encourages interoperability across platforms. As highlighted by ECMA's description, Microsoft's OOXML format would only offer limited interoperability, and even that to a large extent only on Windows. And because Microsoft has 97.5% of the PPA market, exclusively supports OOXML, and does so as the default save format for its PPAs, true competition among alternatives is not possible. Experience has clearly demonstrated Microsoft's ability to exploit its dominant position in the personal computer market to impose de facto industry standards. Equally functional, if not superior, products have been effectively displaced as a result of these practices.

ECIS' Concerns over OOXML

The published OOXML specifications clearly show that the formats were never intended for use by developers wishing to develop PPAs that could run on non-Windows platforms. Instead, the format specifications represent a bloated "dump" of information on how Microsoft Office interacts with the file formats, without intending to disclose information needed for implementing support for OOXML in PPAs running on operating systems other than Windows.

Full file format interoperability support on operating systems such as Linux and Solaris will not, however, be possible, due to Microsoft's deficient disclosure of the OOXML format. Parts of the file format have not been fully disclosed, and some parts of the format rely on dependencies on the Windows operating system and Microsoft Office.

Non-disclosure of parts of the OOXML file format. Despite its length, the OOXML specification approved by ECMA fails to provide enough information for third party PPA developers to implement full support for the OOXML format, particularly on Linux and Solaris operating systems. Of particular concern are parts of the specification that merely specify that an application needs to emulate the behaviour of Microsoft Office, and elements using application-defined values.

- *Elements requiring an application to emulate Microsoft Office:* There are numerous examples of elements in the specification that Microsoft has included in the format for what it calls "compatibility" purposes. These include footnoteLayoutLikeWW8 (Emulate Word 6.x/95/97 Footnote Placement), mwSmallCaps (Emulate Word 5.x for the Macintosh Small Caps Formatting), lineWrapLikeWord6 (Emulate Word 6.0 LineWrapping for East Asian Text), shapeLayoutLikeWW8 (Emulate Word 97 Text Wrapping Around Floating Objects), and useWord2002TableStyleRules (Emulate Word 2002 Table Style Rules). It would have been possible to define the elements in an application-neutral way, but Microsoft decided not to do so. Even if they were required for compatibility purposes, it is difficult to justify not specifying them.

- *Properties depending on application-defined behaviour:* Some elements of the specification are not defined at all. Properties such as the "equationxml" and "gfxdata" contain "application-defined" values which are not disclosed in the specification. Many other elements, such as those dealing with default paragraph properties, rely on application-defined behaviour as default, unless other type of behaviour is defined. Given Microsoft Office's ubiquity, any reference to "application-defined" values in essence means a value defined by Microsoft Office.¹²

Dependencies in OOXML documents on Windows, Office and Microsoft's server-side products. The greatest concern from the point of view of cross-platform interoperability is the presence of Windows and Microsoft Office -dependent content in OOXML documents created in Office 2007. Research conducted by ECIS members has shown that, although the standardised OOXML specifications purport to be platform neutral, Office 2007 in fact stores binary objects and other Windows specific data in OOXML documents. For example, in an OOXML document with embedded content controls, the controls are described in "placeholders" by a Graphical User Interface Descriptor (GUID), which references a control on the Windows platform

These Windows-dependent elements are merely examples, and many others exist. For example Office documents that use macros are not fully usable on other operating systems, due to dependency of macros on the operating system. Similarly, documents protected by Microsoft's Information Rights Management technology can only be opened on fully Window-based environments employing both Microsoft's client and server operating systems.

Research continues to reveal similar elements, which ensure that developers wishing to support OOXML on a non-Windows operating system are faced with a hurdle of trying to reverse-engineer Windows functionality required to support the features present in the documents.

Conclusion

Document interoperability is important for consumers, companies, organisations and governments. Customers should seek to avoid products which create vendor lock in by using undisclosed document formats and creating technical dependencies between the data and the underlying operating system.

ODF as a platform and vendor-independent format shows promise, and its wide adoption would benefit users of all kinds by helping them maintain control of their data and not lock their data into proprietary formats which may someday no longer be supported and readily accessible in a broad range of product offerings. Microsoft's adoption of ODF as its preferred file format would result in the restoration of real competition and consumer choice in PPA applications. Its efforts to supplant ODF with OOXML, on the other hand, will without a doubt make it easier for Windows applications to access data created in Microsoft Office, but it will not provide the level of document interoperability needed to ensure that users' data is owned by users themselves, and not tied to Windows.

¹² Supra, footnote 9.

If Microsoft were seriously committed to competition and choice for consumers, it would truly open up both its binary file formats as well as the new XML formats. Until it does so, competition in the PPA space cannot freely emerge, as Microsoft Office remains the only PPA suite capable of fully supporting Microsoft's document formats.

Further reading

- Groklaw, EOOXML objections¹³
- ODF Alliance, Open XML Fact Sheet¹⁴

¹³ Ibid.

¹⁴ <http://www.odfalliance.org/resources/OfficeOpenXMLFactSheet.pdf>