FSFE Context Briefing

Interoperability woes with MS-OOXML:
Unspecified binary content, Lack of conformance clause, Legal uncertainty

The proposed MS-OOXML/DIS29500 specification raises serious technical and legal concerns.[1] This context briefing highlights three examples of how the proposed specification and its practical implementation in MS Office 2007 hinders interoperability, fosters vendor dependence and results in market distortion.

It does not alleviate concerns that at the ISO Ballot Resolution Meeting for the proposed specification more than 1,000 technical concerns and proposed dispositions required discussion. Participants were only able to discuss between 20 to 30 dispositions and to accept approximately 200 minor editorial corrections in the allocated time. Around 900 dispositions were not discussed.[2]

Example #1: Unspecified binary content in MS Office 2007 generated MS-OOXML files hinders interoperability

Analysis has shown that XLSX documents created by MS Office 2007 have binary content in addition to content described in the proposed MS-OOXML specification. This hinders interoperability and has the potential to reduce document fidelity. The analysis was conducted by downloading a XLSX document from microsoft.com and unpacking the zipped contents to allow review of the internal file structure.[3,4]

The binary content consists of three implementation defined files called printerSettings1.bin, printerSettings2.bin and printerSettings3.bin. They

originate from Microsoft and their content is not described in the proposed specification. Examining the binary files in a HEX editor reveals references to 'Microsoft OneNote Import' and 'Letter'. 'Letter' appears to be a reference to page size.

Referencing page size in a implementation defined binary file is a problematic. Page size information is critical for ensuring the correct layout of a document. European applications without access to the binary information may use A4 page size instead of Letter for displaying the document, thus allowing for more content on each page. Two different users could get the impression they are discussing very different documents when their page numbers do not match.

Example #2: The conformance clause is meaningless

The proposed specification has a loosely worded clause to determine application conformance. It states that a conforming consumer needs to open a conforming document without generating an error and that a conforming producer must be able to create a single conforming document. Any features of the specification

implemented by the applications need to adhere to the definitions in the proposed specification. These terms allow applications that do not even utilise documents to be considered conforming.[5]

An example of an application that adheres to the proposed specification's conformance clause but should

[1] See for example http://www.grokdoc.net/index.php/EOOXML_objections

[2] http://www.consortiuminfo.org/standardsblog/article.php?story=20080229055319727

[3] http://download.microsoft.com/download/a/a/3/aa3411df-5b02-463a-8ab1-9587dd8a2508/Salesdata.xlsx

[4] The zip container contained the following files: ./Content_Types].xml, ./_rels/.rels, ./docProps/app.xml, ./docProps/core.xml, ./docProps/custom.xml, ./xl/_rels/workbook.xml.rels, ./xl/calcChain.xml, ./xl/printerSettings/printerSettings1.bin, ./xl/printerSettings/printerSettings2.bin, ./xl/printerSettings/printerSettings3.bin, ./xl/sharedStrings.xml, ./xl/styles.xml, ./xl/tables/table1.xml, ./xl/theme/theme1.xml, ./xl/workbook.xml, ./xl/worksheets/_rels/sheet1.xml.rels, ./xl/worksheets/_rels/sheet2.xml.rels, ./xl/worksheets/_rels/sheet3.xml.rels, ./xl/worksheets/sheet1.xml, ./xl/worksheets/sheet2.xml, ./xl/worksheets/sheet3.xml

[5] The precise wording states that "A conforming consumer shall not reject any conforming documents of at least one document conformance class. the document type (§4) expected by that application. A conforming producer shall be able to produce conforming documents of at least one document conformance class." (continued overleaf)

not be regarded as conforming is GNU 'cp', a command used to copy files.[6] 'cp' is a technically conforming consumer of the proposed specification because it does not reject a conforming document and any features of the specification implemented in the application (none) are faithful to the specification. 'cp' is also a conforming producer of proposed specification because it can create a conforming document and any features of the specification implemented in the application (none) are faithful to the specification.

The conformance clause for the proposed specification is insufficient because virtually any application that opens or saves files can be considered conforming. The degree to which applications utilise documents is not judged and this allows for misleading claims of specification support.[7] A conformance clause is one of the most important parts of a standard and the text used in DIS29500 is effectively meaningless.

Example #3: Microsoft's Open Specification Promise is not reliable legal coverage for complete interoperability

MS-OOXML files generated by MS Office 2007 contain content that is implementation defined. This is a cause for concern because content not described in the proposed specification has an unclear status regarding coverage under the Microsoft Open Specification Promised (OSP). OSP coverage is limited to patents "that are necessary to implement only the required portions of the Covered Specification that are described in detail and not merely referenced in such Specification." [8]

The OSP states in the final sentence of paragraph two that "No other rights except those expressly stated in

this promise shall be deemed granted, waived or received by implication, exhaustion, estoppel, or otherwise"[9]. It appears reasonable to not rely on the OSP for content necessary to allow interoperability that is not described in detail or referenced in the proposed specification.

This concern becomes more acute if the document is saved in other variations of the proposed specification format. For example, XLSM documents contain unspecified content as well as binary content. XLSB documents contain content stored using a method apparently not described in the proposed specification. XSLX documents with a password are also stored using a document container apparently not covered by the proposed specification.

Conclusion

Binary content, lack of effective conformance clause and legal uncertainty are only a sample of the concerns associated with the proposed MS-OOXML specification.

Given that the ISO process has around 900 unresolved technical comments and will not discuss legal considerations, the suitability of the proposed specification is more than questionable. The only outcome of the proposed specification and its practical implementation in MS Office 2007 is hindered interoperability, vendor dependence and continued market distortion.

In our view there is only one reasonable response by national bodies: move DIS29500 out of the FastTrack process by voting "DISAPPROVE, with comments" and suggest methods of handling the proposed specification through the normal ISO process, ideally by convergence into ISO/IEC 26300, the Open Document Format (ODF).

(continued from previous page) "A conforming application shall treat the information in Office Open XML documents in a manner consistent with the semantic definitions given in this Specification. An application's intended behavior need not require that application to process all of the information in an Office Open XML document. However, the information that it does process shall be processed in a manner that is consistent with the semantic definitions given in this Specification."

[6] Start with a conforming document called 'test.docx'. Run the command 'cp test.docx test_copy.docx' and confirm that no errors are generated and a new copy of the conforming document is created.

[7] Examples of applications named by Microsoft as supporting the proposed specification include: Apple Mac OS X, which has no support for images or embedded objects. iWork, which has no support for spreadsheets, can import only a limited set of documents and has no ability to save documents. Google (Search / Preview), which handles text content but has no support for layout, images or embedded objects.

[8] http://www.microsoft.com/interop/osp/default.mspx

[9] Ibid