

Technical Comments on DIS 29500

Martin Chapman, Consulting Member of Technical Staff, Oracle
Toshihiro Suzuki, Standards Strategy and Architecture Senior Director, Oracle
August 3, 2007

During the review period for the ECMA OOXML specification, a substantial number of issues have been raised that in our opinion cast doubt on the quality and implementability of the specification. Tackling all the issues in a satisfactory manner is not possible in the timeframe permitted and is inappropriate for the Fast Track process.

The prime use case of OOXML is "to be capable of faithfully representing the pre-existing corpus of word processing documents, presentations, and spreadsheets...". Given this use case, it is of utmost importance that the first version of this specification is as future proof as possible. Any changes in the future that break backwards compatibility of documents will violate the prime use case. Therefore time should be spent to get this specification right the first time.

It is Oracle's opinion that given the many technical issues that have been raised during this balloting process, the current specification is not future proof, and hence down the line the prime use case will be violated. There are four areas of concern that must be resolved in order to move forward:

1. Addition of Normative References
2. Removal of VML
3. Removal of all "act like word processor X" Tags
4. Allow Negative Integers for Dates Before 1900

Points 2 and 3 suggest the removal of items from the specification. Rather than delegate to proprietary extensions, it might be acceptable to define a new conformance profile that collects together these legacy and proprietary features. As such, the current application conformance point could be divided into two, where basic application conformance does not require these legacy features, and the additional conformance profile supports basic conformance plus the legacy features.

1. Missing Normative References

Oracle expects a complete list of normative references to include *at least* see XML 1.0, XML Schema (parts 1, 2 and 3), and .Zip File Format Specification in the normative references. Furthermore the version of Zip needs to be defined as per JTC1 directives (the current informative reference effectively points to "the latest version" which can change.) A complete list is included in appendix A.

Under JTC1 directives, any normative references that are not International standards, must either come from an Approved Referencing Organization (ARO), or must be accompanied by a Referencing Explanatory Report (RER). Note that as of January 2007, W3C is an ARO.

2. Removal of Dependencies on Proprietary Technologies

The VML Reference Material (Part 4 Section 6) contains a number of proprietary references. These include Microsoft specific namespaces, references to Windows meta file formats, and dependency on specific versions of Internet Explorer.

On the later, the "target" attribute (Hyperlink Display Target) of several elements has values "_media" and "_search". The description for them is as follows respectively (refer to p4371 of OOXML Part 4, for example);

- "_media" : Specifies that the linked document is loaded into the Media Bar. Available in Microsoft Internet Explorer 6 or later.
- "_search": Specifies that the linked document is loaded into the browser's search pane. Available in Microsoft Internet Explorer 5 or greater.

Such explicit proprietary dependencies have no place in an international standard therefore consideration should be given to remove section 6.

3. Under-Defined Attributes

Throughout Part 4, Section 2.15 a number of compatibility setting are defined: useWord2002TableStyleRules, useWord97LineBreakRules, wpJustification, autoSpaceLikeWord95, footnoteLayoutLikeWW8, lineWrapLikeWord6, mwSmallCaps, shapeLayoutLikeWW8, suppressTopSpacingWP, truncateFontHeightsLikeWP6, useWord2002TableStyleRules.

The semantics of these tags are not defined and thus it is not possible to implement these without knowledge that exists outside of the specification. This is not acceptable for an international standard.

Furthermore, given the possible complexity of implementing these tags, there is no option to ignore the tags. In other words if they are present in a document the application must behave according to the tag (which of course is not defined).

At the very least text should be added to say that applications may ignore these tags but must preserve them in documents if they are present (i.e. must not remove them).

Either the semantics of these tags must be defined, or they must be removed from the specification.

4. Allow Negative Dates Before 1900

As an international standard there is no good reason to propagate a flawed artifact of a legacy proprietary products. While a minor technical change, the ability to represent a full range of date values prevents unanticipated problems from occurring when date values before 1900 are necessary.

Appendix A: Missing Normative References

These are the references from each part's bibliography that need to be defined in the relevant Normative References. All ISO references and W3C and Unicode Consortium references can be directly moved - the later two coming from Approved Referencing Organizations (ARO)

In addition, RER descriptions should be prepared for specifications originated by IANA, IETF and Dublin Core and they should be moved to Normative References. Similarly, specifications such as PANOSE by Hewlett Packard, ZIP format by PKWARE, Inc are one of the few vendor-dependant specifications. These are necessary in terms of OOXML implementation and should be included in Normative References with RER description.

Not included are wmf and emf, which would be needed as Normative References should VML remain.

Note:

Tags of Dublin Core are used in Core Properties (See " 10.1 Core Properties Part " in "Part 2: Open Packaging Conventions"). PANOSE can be specified as Font Substitution Data of WordprocessingML (eg. `<w:panose1 w:val="020B0A04020102020204" />`). (See "2.8.2. 13 panose1 (Pansose-1 Typeface Classification Number) " in " Part 4: Markup Language Reference").

ZIP format in OOXML specification is explained in "physical mapping to a ZIP archive" and used in Office products. (See "9.2 Mapping to a ZIP Archive" in " Part 2: Open Packaging Conventions "). Note a precise version for ZIP needs to be referenced (the part 2 reference does this, the part 1 one does not.)

<< Part 1: Fundamentals >>

Annex A. Bibliography

[IANA]

- Character Sets from IANA, as specified at <http://www.iana.org/assignments/character-sets>

[IETF]

- RFC 2119, Bradner, Scott, 1997: "Key words for use in RFCs to Indicate Requirement Levels."
<http://www.ietf.org/rfc/rfc2119.txt>.

- RFC 2045, Borenstein, N., and N. Freed. "Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies.
" The Internet Society. 1996. <http://www.rfc-editor.org>
- RFC 2616, Berners-Lee, T., R. Fielding, H. Frystyk, J. Gettys, P. Leach, L. Masinter, and J. Mogul. "Hypertext Transfer Protocol-HTTP/1.1." The Internet Society. 1999. <http://www.rfc-editor.org>
- RFC 3066, Alvestrand, H. "Tags for the Identification of Languages." The Internet Society. 2001. <http://www.rfc-editor.org>
- RFC 3339, Klyne, G. and C. Newman "Date and Time on the Internet: Timestamps." The Internet Society. 2002.
<http://www.rfc-editor.org>
- RFC 3629, Yergeau, F. "UTF-8, a transformation format of ISO 10646." The Internet Society. 2003.
<http://www.rfc-editor.org>
- RFC 3986, Berners-Lee, T., R. Fielding, and L. Masinter. "Uniform Resource Identifier (URI): Generic Syntax." The Internet Society. 2005. <http://www.rfc-editor.org>

[Unicode Consortium]

- The Unicode Consortium. The Unicode Standard, Version 4.0, defined by: The Unicode Standard, Version 4.0 (Reading, MA, Addison-Wesley, 2003. ISBN 0-321-18578-1).

[W3C]

- XML, Bray, Tim, Eve Maler, Jean Paoli, C. M. Sperberg-McQueen, and François Yergeau (editors). "Extensible Markup Language (XML) 1.0," Third Edition. World Wide Web Consortium. 2004. <http://www.w3.org/TR/2004/REC-xml-20040204/xml11-20040204/>
- XML Base, Marsh, Jonathan. "XML Base." World Wide Web Consortium. 2001.
<http://www.w3.org/TR/2001/REC-xmlbase-20010627/>
- XML Namespaces, Bray, Tim, Dave Hollander, Andrew Layman, and Richard Tobin (editors). "Namespaces in XML 1.1." World Wide Web Consortium. 2004. <http://www.w3.org/TR/2004/REC-xml-names11-20040204/>
- XML Path Language Specification, Version 1.0, W3C Recommendation 16 November 1999
<http://www.w3.org/TR/xpath>
- XML Schema Part 0: Primer Second Edition, W3C Recommendation 28 October 2004
<http://www.w3.org/TR/xmlschema-0/>

- XML Schema Part 1: Structures Second Edition, W3C Recommendation 28 October 2004
<http://www.w3.org/TR/xmlschema-1/>

- XML Schema Part 2: Datatypes Second Edition, W3C Recommendation 28 October 2004
<http://www.w3.org/TR/xmlschema-2/>

[Vender Specification]

- PANOSE Classification Guide, Version 1.2, Hewlett Packard Co., 1992. (Hewlett Packard)

- ZIP File Format Specification from PKWARE, Inc., as specified in appnote, the Application Note on the Zip file format, at

<http://www.pkware.com.>(PKWARE, Inc)

<< Part 2: Open Packaging Conventions >>

Annex I. Bibliography

[ISO]

- ISO/IEC Directives Part 2, Rules for the structure and drafting of International Standards, Fourth edition, 2001, ISBN 92-67-01070-0.

[Unicode Consortium]

- The Unicode Standard, Version 3.0, by the Unicode Consortium; Addison-Wesley Publishing Co, ISBN 0-201-8 61633-5, February 2000.

The latest version can be found at the Unicode Consortium's web site, www.unicode.org, at this writing.

[Dubin Core]

- Dublin Core Element Set v1.1. <http://purl.org/dc/elements/1.1/>

- Dublin Core Terms Namespace. <http://purl.org/dc/terms/>

[W3C]

- Extensible Markup Language (XML) 1.0 (Third Edition), W3C Recommendation, 04 February 2004.

- Namespaces in XML 1.1, W3C Recommendation, 4 February 2004.

- W3C NOTE 19980827, Date and Time Formats, Wicksteed, Charles, and Misha Wolf, 1997,
<http://www.w3.org/TR/1998/NOTE-datetime-19980827>.

- XML Base, W3C Recommendation, 27 June 2001.

- XML Path Language (XPath), Version 1.0, W3C Recommendation, 16 November 1999.

- XML Schema Part 1: Structures, W3C Recommendation, 28 October 2004.

- XML Schema Part 2: Datatypes, W3C Recommendation, 28 October 2004.
- XML-Signature Syntax and Processing, W3C Recommendation, 12 February 2002.

[IETF]

- RFC 2616 Hypertext Transfer Protocol-HTTP/1.1, The Internet Society, Berners-Lee, T., R. Fielding, H. Frystyk, J. Gettys, P. Leach, L. Masinter, and J. Mogul, 1999, <http://www.rfc-editor.org>.
- RFC 3986 Uniform Resource Identifier (URI): Generic Syntax, The Internet Society, Berners-Lee, T., R. Fielding, and L. Masinter, 2005, <http://www.rfc-editor.org>.
- RFC 3987 Internationalized Resource Identifiers (IRIs), The Internet Society, Duerst, M. and M. Suignard, 2005, <http://www.rfc-editor.org>.
- RFC 4234 Augmented BNF for Syntax Specifications: ABNF, The Internet Society, Crocker, D., (editor), 2005, 21 <http://www.rfc-editor.org>.

[Vender Specification]

- ZIP File Format Specification, Version 6.2.1, PKWARE Inc., 2005. (PKWARE, Inc)

<< Part 3: Primer >>

Bibliography N/A

<< Part 4: Markup Language Reference >>

Bibliography N/A

<< Part 5: Markup Compatibility and Extensibility >>

Annex B. Bibliography

[ISO]

- ISO/IEC 197575-4, Information technology - Document Schema Definition Languages (DSDL) - Part 4: Namespace-based Validation Dispatching Language (NVDL).

- ISO/IEC Directives Part 2, Rules for the structure and drafting of International Standards, Fourth edition, 2001, ISBN 92-67-01070-0.

[W3C]

- Extensible Markup Language (XML) 1.0 (Fourth Edition), W3C Recommendation, 16 August 2006.
- Namespaces in XML 1.0 (Second Edition), W3C Recommendation, 16 August 2006.
- XML Base, W3C Recommendation, 27 June 2001.
- XML Schema Part 1: Structures Second Edition, W3C Recommendation, 28 October 2004.
- XML Schema Part 2: Datatypes Second Edition, W3C Recommendation, 28 October 2004.