



Software Libre

A r g e n t i n a

www.solar.org.ar

info@ solar.org.ar

Buenos Aires, miércoles 8 de Agosto de 2007

Ing. José F. López

Director General del IRAM

(Instituto Argentino de Normalización y Certificación)

S / D

De nuestra mayor consideración:

Por la presente nos dirigimos a Ud. para solicitar que IRAM se sume a otros miembros de la ISO, tales como Brasil, México y Chile, y tome participación en el proceso de estandarización por vía rápida ("fast-track") del documento DIS 29500 (ECMA 376-Office Open XML Document format, o OOXML [1]) que actualmente se encuentra en proceso de revisión en la ISO (en el período de "ballot" de 5 meses a finalizar el 2/9/2007) y asimismo para requerirle que IRAM formule un voto *negativo* a dicha propuesta.

A continuación exponemos los motivos de este pedido:

OOXML es una especificación de más de 6000 páginas [2], generada por Microsoft para su nuevo producto Office 2007, enviada en Diciembre/2005 a ECMA [3] para su aprobación como estándar, y aprobada por ECMA en menos de un año (a un ritmo sin precedentes de revisión/edición/aprobación de 18.3 páginas por día y sin visibilidad pública del proceso) y finalmente enviada por ECMA a la ISO solicitando su aprobación como estándar internacional ISO por la vía fast-track.

Dada su complejidad, la multiplicidad de dependencias con un solo producto informático, y los graves problemas detectados en la especificación (más adelante expuestos), la promoción de ésta especificación a estándar ISO resultará en beneficio y utilidad de una sola empresa (Microsoft, que comercializa el único producto que podrá implementar la especificación en un 100%) y marginalmente y en muy menor medida a empresas "partners" que logren proveer algún servicio implementando alguna parte del formato.

Quedarán afuera de los beneficios las demás partes interesadas, esto es, el resto de las empresas, administraciones públicas, ciudadanos, trabajadores informáticos, usuarios finales, o cualquier otra persona o entidad que quiera utilizar las ventajas de un formato electrónico abierto y estándar *sin depender* de un proveedor particular de software o que *demande* un formato electrónico estándar que sea interoperable en ambientes con plataformas informáticas heterogéneas (PC, Macintosh, Mainframes, Windows, Linux, Unix, etc.).

Por ejemplo, si en Argentina una empresa local productora de software de Oficina quisiera desarrollar un procesador de texto o planilla de cálculo que soporte la edición y generación de archivos en formato "ISO" OOXML le resultaría prácticamente imposible lograrlo, ya que aún cuando podría acceder a +6000 páginas que describen los elementos y atributos XML que conforman el formato,

tendría muy pocas pistas acerca de la "semántica" de ese XML. Ejemplo: esta empresa sabría que en un archivo "ISO" OOXML puede encontrar elementos como "useWord2002TableStyleRules" , "useWord97LineBreakRules", "AutoFill" y/o "lineWrapLikeWord6" pero no sabría cómo procesarlos para mostrarlos en pantalla ya que OOXML no da ninguna información sobre estos elementos (ejemplos como estos abundan en todo OOXML) [4].

Es nuestro entender que un proceso de estandarización no debe perseguir un fin tan limitado, sino que, tal cual lo expresa IRAM, *"la normalización implica la participación de personas que representan a distintas organizaciones de los tres sectores involucrados: productores, consumidores e intereses generales. Estos representantes aportan su experiencia y sus conocimientos para establecer soluciones a problemas reales o potenciales"*. [5][6]

Cada vez más, las administraciones públicas de los países del mundo están adoptando políticas para asegurar el acceso irrestricto a documentos electrónicos, facilitar la interoperabilidad interna y externa, y garantizar la preservación de la información electrónica de interés público a corto, mediano y largo plazo (en el orden de los 1 año hasta los 500 años). [7]

Por lo tanto, las decisiones adoptadas por cada NB (miembro nacional o "national body" de la ISO) para promover (o no) especificaciones de formatos electrónicos al status de estándar internacional ISO tendrán consecuencias directas e importantes en este proceso.

Por todo lo expuesto, resulta clave que cada NB, sea miembro "P" o miembro "O" de ISO (como lo es IRAM), no mantenga un rol pasivo sino participativo en estas cuestiones, y forme una comisión para analizar si OOXML cumple con los requisitos para convertirse en un estándar ISO y exprese su voto en tiempo y forma previo a la resolución final prevista para el 2 de Setiembre del presente año.

En el caso de que se otorgue por la vía fast-track el sello de "estándar ISO" a OOXML, sin el debido análisis de cada país miembro o, peor aún, por presión de intereses comerciales, a nuestro entender se estará devaluando el prestigio de ISO como organización productora de estándares internacionales de calidad y se perderá la oportunidad de lograr la verdadera apertura de los formatos de documentos electrónicos de oficina, para beneficio de todas las partes interesadas (empresas, administraciones públicas, ciudadanos, trabajadores informáticos, usuarios finales).

Seguidamente se enumeran las razones por las cuales IRAM debería tomar cartas en este asunto y votar *negativamente* a la estandarización de OOXML vía fast-track, solicitando que este estándar sea considerado por la vía normal prevista por la ISO en sus procedimientos:

OOXML no cumple con los requisitos para calificar como estándar internacional, y mucho menos mediante la vía fast-track:

Adjunto a la presente se incluye un informe titulado "[El caso contra OOXML](#)" donde se discute por qué el documento DIS 29500, "Office Open XML" (OOXML), no cumple con los siguientes criterios definidos por ISO y otras organizaciones para una norma internacional:

- Precisión, repetibilidad, generalidad
- Parámetro de excelencia, mejores prácticas consolidadas
- Interoperabilidad y capacidad de transferencia
- Adaptación cultural y lingüística

Se han detectado y alertado contradicciones durante el período de 30 días previsto por la ISO en su procedimiento de estandarización fast-track. Dichas contradicciones no han sido todavía resueltas y el tiempo previsto en el procedimiento fast-track no es el adecuado para hacerlo, dada la gravedad e importancia de las mismas:

El procedimiento de fast-track estipulado por ISO/IEC JTC 1 establece un período de revisión de 30 días del estándar propuesto, para que los NB identifiquen cualquier contradicción percibida con otro estándar JTC 1, ISO o IEC [8].

El grupo de documentos distribuidos por el JTC 1 de la ISO el miércoles 28 de Febrero de 2007 que contiene las respuestas de los diferentes NB cumplido el período de 30 días de revisión indica que 6 (seis) países han presentado una desaprobación formal a la consideración de este estándar mediante fast-track y otros 5 (cinco) países han expresado serias preocupaciones [9]. En particular, estos son extractos de algunas de las respuestas:

- Alemania (DIN):
"El NB de Alemania prefiere tener estándares armonizados a tener dos estándares que satisfagan básicamente los mismos requerimientos de usuarios. Por lo tanto, el NB de Alemania solicita que JTC1 debe, previamente al inicio del período de 5 meses de ballot:
 - *solicitar una lista de las funcionalidades que potencialmente no están previstas en el estándar ISO/IEC 26300 para asegurar la compatibilidad con los formatos binarios previos de Microsoft.*
 - *permitir que OASIS, como desarrollador de ISO/IEC 26300 analice si dicho estándar puede ser razonablemente mejorado para que tales requerimientos de compatibilidad puedan ser alcanzados.*

Solamente una vez que dichos análisis demuestren ser negativos, debe continuarse con el procesamiento de DIS 29500.

- Inglaterra (BSI):
"Inglaterra considera que existe una contradicción entre ECMA 376 (DIS 29600) y estándares ISO/IEC existentes, en particular, ISO/IEC JTC1 26300 "Open document format for Office Applications..."
"Inglaterra considera que la sección 13.2 [de las directivas del JTC 1] no puede ser satisfecha y que por lo tanto este documento no puede ser propuesto para el ballot de 5 meses previsto en el procedimiento de fast track..."
"Dada la complejidad y el tamaño de este documento, Inglaterra no cree poder cumplir con su obligación de revisar esta especificación y mantener la integridad del proceso y la reputación del JTC1 dentro de los cinco meses del período de ballot..."
"Inglaterra estaría de acuerdo en que ECMA 376 sea propuesto como un 'Borrador de Comité' junto con una 'Propuesta de nuevo ítem de trabajo', de tal forma que se siga el procedimiento normal en cambio del procedimiento 'fast-track'.
- Nueva Zelanda (SNZ):
"[Lo dicho] indica que N8455 [Office Open XML] coincide con las especificaciones ODF existentes, y el tener dos estándares coincidentes que resuelven el mismo problema resultará en confusión para los usuarios y falta de interoperabilidad. Esta coincidencia puede ser considerada una violación de la Resolución 28,2000, y por lo tanto N8455 debe ser dado de baja de este procedimiento fast-track.
- Republica Checa (CNI):
"Los documentos abiertos, y en general los estándares abiertos, son muy importantes para el intercambio global de información y por lo tanto se necesita una amplia discusión de todas las partes interesadas... " "La República Checa sugiere que se utilice el procedimiento estándar para el desarrollo de un estándar ISO/IEC a partir del ECMA-376."

- Australia(SA):

"Australia considera que existen dificultades en conducir propiamente una revisión completa de un documento de más de 6000 páginas en un mes...." "Australia propone que este documento sea devuelto a SC34 para su discusión previamente a ser enviado a un ballot DIS, dada la cantidad significativa de cuestiones que deben ser clarificadas".
- Finlandia (SFS):

"Considerando el tamaño, complejidad y alcances de esta propuesta de ECMA, debemos alertar algunas preocupaciones sobre el procedimiento a seguir."

"Si se considera la rapidez del proceso de ECMA, la rapidez del procedimiento fast-track, el tamaño (más de 6000 páginas) y la complejidad de la especificación propuesta, tenemos serias dudas de que este o cualquier otro NSB [miembro nacional de la ISO] pueda cumplir exitosamente con sus obligaciones de revisar esta especificación y mantener la integridad del proceso y la reputación de JTC1."

"Esta especificación contiene a su vez especificaciones completas de dos lenguajes gráficos de vectores diferentes (VML y DrawingML), una especificación completa para la representación de ecuaciones matemáticas (OOMML), una especificación completa de un lenguaje de evolución de esquemas XML (Markup Compatability ML) y un lenguaje completo de citas bibliográficas, además de otros. De estándares similares producidos por la W3C, tales como SVG o MathML, sabemos que el desarrollo y la revisión de por lo menos uno de estas sub-especificaciones requeriría de un grupo experto de 2-3 años de trabajo. Pero ECMA, en un proceso que no tuvo mucha visibilidad para el público, produjo una especificación que incluye todo eso, y el ciclo de revisión y aprobación les tomo menos de un año."

"Por lo tanto, consideramos que lo mejor es que Office Open XML sea removido del procedimiento de fast-track del JTC1 y que, o se lo envíe a un grupo de trabajo para una revisión más profunda, dividido en subsecciones de tamaño razonable, por ejemplo 500 páginas, para su aprobación normal, o (preferiblemente) que Office Open XML sea armonizado con el estándar existente ISO/IEC 26300 'Open Document Format'."

Existe amplia evidencia de que ECMA, la organización que propuso Office Open XML, ha hecho un trabajo editorial pobrísimo:

- no ha revisado errores ortográficos:

La especificación esta plagada de ellos. Una revisión realizada personalmente por el que suscribe detectó más de 100 errores ortográficos de todo tipo, incluyendo frases completas incluidas en lugares extraños, sin ninguna relación al texto (ej: Parte 2, última página, último párrafo). Un extracto de los errores mencionados se acompaña a la presente ("[Anexo 2: Extracto errores ortográficos en ECMA-376-OOXML](#)").

- no ha validado los elementos, atributos y ejemplos XML: en todo DIS 29500 (OOXML) es común que los elementos de los distintos esquemas se refieran con nombres erróneos, en secciones distintas de la misma parte, en partes distintas o entre el texto de las partes y los esquemas XML provistos por ECMA como anexos a OOXML. Ejemplo:

La sección 2.15.1.10 nombra en 6 líneas consecutivas de la misma pagina (pagina 1118) el elemento "Automatically Hyphenate Document Contents When Displayed" de tres maneras distintas:

"autoHypehenation" en la línea 18

"autoHypehpnation" en la línea 20

"autoHyphenation" en la línea 14

Un extracto de este tipo de errores se acompaña a la presente ("[Anexo 3: Extracto de errores en referencias XML ECMA-376-OOXML](#)").

- los ejemplos XML están duplicados, son confusos y con muchos errores:

Acerca de este problema, el grupo de expertos del NB de Inglaterra que está analizando OOXML expresa: *"Hay aproximadamente 2300 ejemplos en la sección WordProcessingML de la especificación. Dichos ejemplos fueron testeados para determinar si son bien-formados y si validan contra su esquema, utilizando software customizado. Aproximadamente 300 ejemplos, más del 10%, tienen errores..."* *"Opinión: aunque la existencia de algún error de este tipo es entendible en una especificación tan grande, el notable volumen de errores detectado indica que esta especificación no ha sido objeto de una revisión técnica rigurosa previo a convertirse en un estándar ECMA, y por lo tanto no sería viable su estandarización mediante proceso de fast-track"*. [10]

- no hay un criterio uniforme para identificar partes normativas o informativas en la especificación:

Un extracto de algunos ejemplos de esta problemática recopilado por el que suscribe se acompaña a la presente ("[Anexo 4 - Ejemplos ambigüedades normativas/informativas en ECMA-376-OOXML](#)").

Comentando sobre esta cuestión, un miembro del NB de USA, Jon Bosak [11], ni más ni menos que uno de los creadores del estándar XML, expresa: *"Cuanto más avanzo con esto [la revisión de OOXML], me siento más inclinado en coincidir con la opinión de que 'si [OOXML] hubiera ingresado por el procedimiento normal de la ISO, yo diría que se encuentra en un estado de 'borrador de trabajo' y ni siquiera se encuentra listo para registrarlo como un 'borrador de comité' y asignarle un número'"* [12]

El grupo experto del NB de Inglaterra (BSI, que tradicionalmente hace los mejores análisis y comentarios al estudiar este tipo de estándares) expresó: *"[El gran número de errores editoriales] sugiere una falta de rigor en el proceso de preparación de este documento, y ha sido general [en el grupo] la expresión de sorpresa ante el hecho de que ECMA le haya puesto su firma a un texto tan descuidado."* [13]

Alex Brown, el coordinador ese mismo grupo de BSI-Inglaterra, expresó: *"Aunque el voto de Inglaterra no está decidido, y no tendría sentido tratar de adivinarlo, personalmente creo que una postura no se adoptará: el sí incondicional. Esto es así porque , dejando de lado consideraciones más amplias, el texto está **muy pobremente escrito**. Mi colega, y panelista en BSI, Andrew Sales, me separó este párrafo de la página 1912 de la especificación de SpreadsheetML como un ejemplo demostrativo: [ejemplo obviado]*

*Tal "parafernalia" **no tiene lugar en un estándar ISO**. Viendo esto, uno solo puede preguntarse acerca del rigor de ECMA en sus procesos de estandarización y concluir que **algo está funcionando mal y que debe ser corregido**"* [14]

Uno de los presupuestos del procedimiento de fast-track es que el estándar propuesto ya ha sido rigurosamente revisado y corregido por la organización que lo propone, y que, por lo tanto no requerirá que los NB dupliquen este trabajo. Todo indica que ECMA no ha hecho este trabajo y surge la duda de cuáles y cuántos miembros del grupo de trabajo de ECMA (TC-45) encargado de editar, revisar y aprobar este estándar ha leído realmente la especificación enviada a ISO.

Los cinco meses previstos por el procedimiento de fast-track no son suficientes para evaluar los méritos técnicos de OOXML para convertirse en un estándar internacional ISO

Este problema ha sido y viene siendo alertado por diferentes NB:

- 8 de los 20 de los países miembros que respondieron al "envío de contradicciones" solicitado por la ISO como parte del procedimiento fast-track, expresaron que el tiempo previsto según el procedimiento fast-track no era el adecuado para una especificación de este tamaño (más de 6000 páginas).
- El grupo de expertos del NB de Inglaterra (BSI) que está analizando OOXML expresó que *"[E]s común la opinión de que aún dado el panel de 30 personas, hay insuficiente tiempo para hacerle justicia al texto de DIS 29500 y proveer el nivel de escrutinio que idealmente debería tener"*. [15]

Un ejemplo del tipo de problemas que no puede ser analizado y escrutado suficientemente es el relacionado a la reutilización de estándares existentes. En general, al desarrollar un nuevo estándar se intenta siempre reusar estándares existentes de tal forma de no estar "inventando la rueda" cada rato. Por lo tanto, si un formato de documento involucra ecuaciones matemáticas, y ya existe un estándar para representar ecuaciones matemáticas, entonces uno debe re-usarlo y no inventar una variante propia del mismo.

En lo relativo a OOXML, el estándar existente para ecuaciones matemáticas es MathML, establecido por la W3C [16]. ECMA-376 (OOXML) no reusa dicho estándar, sino que define uno nuevo (en 120 páginas).

De la misma forma, ya existe un estándar desarrollado por la W3C para representación de gráficos vectoriales y se denomina SVG, el cual es soportado por varios productos de software. Sin embargo, OOXML define dos nuevos formatos para el mismo fin: VML (especificación de 617 páginas) y DrawingML (1131 páginas), y ninguno de los dos reusa SVG.

Parecería claro que VML debería ser eliminado de OOXML y quizás DrawingML también. Si OOXML hubiera ingresado a ISO para una revisión "normal" y no por vía fast-track, entonces habría posibilidad de que el comité apropiado de JTC1 analice y discuta en detalle la necesidad y las justificaciones de ECMA para apartarse de estándares existentes.

Solicitamos que IRAM tenga a bien considerar todos los hechos y motivos expuestos y que, como organización que promueve la normalización en beneficio de consumidores, de las propias empresas y de la sociedad en general, se pronuncie negativamente a la estandarización vía fast-track de ECMA-376 Office Open XML (ISO DIS 29500), requiriendo que dicha especificación sea re-enviada por ECMA para su tratamiento por la vía normal.

Referencias:

[1] http://www.ecma-international.org/news/TC45_current_work/TC45_available_docs.htm

[2] Un colaborador de CNI (el Instituto de Estandarización de República Checa) se tomó el trabajo de imprimir y encuadernar las +6000 páginas: http://blog.janik.cz/archives/2007/05/19/T20_32_07/

[3] <http://www.ecma-international.org/>

[4]

Ejemplos de elementos que la especificación solamente nombra, y que no provee ninguna definición sobre cómo implementarlos.

AutoFill element: Parte 4, Section 6.4.2.4

AutoLine element: Parte 4, Section 6.4.2.5

Camera (Camera Tool): Parte 4, Section 6.4.2.8

autoSpaceLikeWord95 element: Parte 4, Section 2.15.3.6 [page 1378]

footnoteLayoutLikeWW8 element: Parte 4, Section 2.15.3.26 [page 1416]

lineWrapLikeWord6 element: Parte 4, Section 2.15.3.31 [page p1426]

useWord2002TableStyleRules element: Parte 4, Section 2.15.3.63 [page 1481]

useWord97LineBreakRules element: Parte 4, Section 2.15.3.64 [page p1482]

etc.

[5] Sitio Web IRAM - Qué es la normalización?:

<http://www.iram.com.ar/normalizacion/normalizacion.htm#que>

[6] Con respecto a las "partes interesadas", viene al caso mencionar que distintos NB han solicitado comentarios al público en general respecto a la conveniencia de estandarizar OOXML y han recibido gran cantidad de comentarios desfavorables. Ejemplo: de 250 respuestas recibidas hasta ahora por SCC, el organismo de estandarización de Canadá, el 98% son desfavorables a la estandarización de OOXML:

<https://forums.scc.ca/forums/scc/dispatch.cgi/public/docProfile/100009/d20070504104953/No/t100009.htm>

[7] Ejemplos:

- Estado de Nueva York-USA: <http://www.assembly.state.ny.us/leg/?bn=A08961&sh=t>
- Estado de Texas-USA:
<http://www.capitol.state.tx.us/BillLookup/History.aspx?LegSess=80R&Bill=SB446>
- Estado de Minnesota-USA:
<http://www.revisor.leg.state.mn.us/bin/bldbill.php?bill=H0176.0.html&session=ls85>
- Gobierno de Brasil: <https://www.governoeletronico.gov.br/acoes-e-projetos/e-ping-padroes-de-interoperabilidade>
- Europa: <http://ec.europa.eu/idabc/en/document/3439>
- Hong Kong: <http://www.ogcio.gov.hk/eng/infra/download/s18.pdf>
- Dinamarca: <http://ec.europa.eu/idabc/en/document/6768/469>
- Gobierno de Misiones-Argentina:
http://www.ccpm.misiones.gov.ar/index.php?option=com_content&task=view&id=34&Itemid=38

[8] ISO/IEC JTC 1 Directives, 5th Edition, Version 2.0. iso. Versión obtenida el 2007-01-28:

<http://www.jtc1sc34.org/repository/0725c.htm>

[9] Las respuestas de los NB están disponibles en el sitio web del JTC 1 de la ISO en el siguiente

archivo: <http://jtc1sc32.org/doc/recent/JTC001-N-8530.zip>

[10] http://xmlopen.org/ooxml-wiki/index.php/2._WordprocessingML_Reference_Material#Errors_in_examples

[11] http://en.wikipedia.org/wiki/Jon_Bosak

[12] <http://www.ibiblio.org/bosak/v1mail/200706/2007Jun22-081045.eml>

[13] http://xmlopen.org/ooxml-wiki/index.php/Convenor's_Interim_Report#General_Editorial

[14] "OOXML: The End Game (and Beyond) ", Sección "The UK Vote", segundo párrafo:

<http://www.adjb.net/index.php?entry=entry070708-121943>

[15] http://xmlopen.org/ooxml-wiki/index.php/Convenor's_Interim_Report#General_Progress

[6] <http://www.w3.org/>

Anexo 1 - El caso contra OOXML

adjuntar el texto "El caso contra OOXML":

<http://people.ffii.org/~abarrio/openxml.info/ElCasoContraOOXML-20070630opentia.pdf>

Anexo 2 - Extracto errores ortográficos en ECMA-376-OOXML

Part 4, Section 2.3.2.5 color (Run Content Color): "*Te resulting themeShade value in the file format would be 66. end example]*"

Part 3, Section 5.8.4 Fills: "*The se types describe the general structure of all fills;*"

Part 1, Section 12.3.6 Custom XML Mappings Part: "*Single Cell Table Defintions (§12.3.19) parts.*"

Part 2, Section 12.2.4.26 Relationships Transform Algorithm: "*that do not have eitheran*"

Part 3, Section 2.8.5 Linked Styles: "*must referrence the paragraph version*"

Part 3, Section 2.10.4 Numbering Definition Instances: "*the abtract numbering definition.*"

Part 3, Section 3.1.4 fileVersion: "*user to initate the action.*"

Part 3, Section 3.16.4 Cell References: "*and a positive offset indicateing a row*"

Part 3, Section 4.5.2 Slide Update Info: "*The part is indentified for each slide by a relationship with the following characteristics*"

Part 3, Section 5.3.2.17 Text Properties: "*The compelex type CT_TableStyleTextStyle*"

Part 3, Section 5.15.5.1.3 Style Definition: "*the index defined in this complex typem is used*"

Part 3, Section 5.15.6.1.4 Child Order Type: "*The simple type ST_ChildOrderType specifices*"

Part 3, Section 5.15.6.1.8 Parameter ID: "*unknown – An unkown parameter id.*"

Part 4, Section 2.3.1.5 between (Paragraph Border Between Identical Paragraphs): "*...If those values*

were identical, then paragraph one would have a border below it, and paragraph two would have no top border. end example]

Part 4, Section 2.3.3.20 pgNum (Page Number Block): "This is the current page: in the document"

Part 4, Section 2.3.3.2 control (Floating Embedded Control): "the document shall be considered non-conformant"

Part 4, Section 2.4.43 tblCellSpacing (Table Cell Spacing Default): "This table cell spacing default is specified using the following WordprocessingML:"

Part 4, Section 2.6.11 pgMar (Page Margins): "one-half of an inch from the page extents"

Part 4, Section 2.11.19 numRestart (Footnote and Endnote 1 Numbering Restart Location): "each page or section (a value of continuous)."

Part 4, Section 2.13.4.2 comment (Comment Content): "The initials attribute specifies that the initials"

Part 4, Section 2.15.1.16 caption (Single Caption Type Definition): "below the object it is labelling"

Part 4, Section 2.15.1.17 1 captions (Caption Settings): "added by specifying the chapNum"

Part 4, 2.18.4 ST_Border (Border Styles): "Speci Specifies an art border consisting"

Part 4, 3.2.3 customWorkbookView (Custom Workbook View): "Correponds to a sheetId of a sheet in the sheets collection."

Part 4, 3.3.1.31 dataValidations (Data Validations): "Data validation is used to specify constraints on the type of data"

Part 4, 3.3.1.37 f (Formula): "an array formula shall be calcuated"

Part 4, 3.8.31 numFmts (Number Formats): "When laoding in CHT (Tawian only)"

Part 4, 3.8.31 numFmts (Number Formats): "Hijri calander"

Part 4, Section 3.8.31 numFmts (Number Formats): "Language info is a 32-bit value entered in hexidecimal format."

Anexo 3 - Extracto de errores en referencias XML ECMA-376 OOXML

Comment 1

Part 3, Section 5.15.6.1.32 Text Anchor Horizontal reads:

"The simple type ST_AnchorHorizontal is responsible for anchoring text horizontally"

No reference of this type was found in Part 4 nor in Part 3.

Proposed change: if this element is missing in Part 4 then provide it, if not, then correct the text in Part 3 and/or in Part 4 to reference the proper type name.

Comment 2

Part 3, Section 5.3.2.22 Table Style:

There are three references to what it seems to be the same element but with distinct names in this normative section:

"band2Vertical" in line 5 of page 277

"band2Vertial" in line 6 of page 277

"band2V" in line 34 of page 277

The same goes to the "band1Vertical", "band2Horizontal" and "band1Horizontal" elements

Proposed change: if this are distinct elements, then define them. Otherwise, correct the elements names.

Comment 3

Part 3, Section 5.15.6.1.33 Text Anchor Vertical reads:

"The simple type ST_AnchorVertical is responsible for anchoring the text vertically."

No reference of this type was found in Part 4 markup language reference nor in Part 3.

Proposed change: if this element is missing in Part 4 then provide it, if not, then correct the text in Part 3 and/or in Part 4 to reference the proper type name.

Comment 4

Part 3, Section 5.15.6.1.8 Parameter ID reads:

"bkPtFixedVal – specifies where the sname should break if bkpt is set to fixed"

No definition of the term "sname" was found in part 3.

Proposed change: define "sname" or correct the text.

Comment 5

Part 4, Section 2.11.17 numFmt (Footnote Numbering Format) reads:

"This element specifies the numbering format which shall be used to determine the footnote or endnote reference mark value for all automatically numbered footnote and endnote reference marks (those without the suppressRef attribute set)."

The definition of the suppressRef attribute was not found in Part 4 nor in any schema.

Proposed change: define the suppressRef attribute or correct the text, and update the schema accordingly.

Comment 6

Part 4, 2.11.18 numFmt (Endnote Numbering Format) reads:

"This element specifies the numbering format which shall be used to determine the footnote or endnote reference mark value for all automatically numbered footnote and endnote reference marks (those without the suppressRef attribute set)."

The definition of the suppressRef attribute was not found in Part 4 nor in any schema.

Proposed change: define the suppressRef attribute or correct the text, and update the schema accordingly.

Comment 7

Part 4, Section 2.3.1.41 textDirection (Paragraph Text Flow Direction) reads

"...This element specifies the direction of the text flow for this paragraph.

If this element is omitted on a given paragraph, its value is determined by the setting previously set at any level of the style hierarchy (i.e. that previous setting remains unchanged). If this setting is never specified in the style hierarchy, then the paragraph shall inherit the text flow settings from the parent section.

[Example: Consider a document with a paragraph in which text should flow bottom to top vertically, and left to right horizontally. This setting would be specified with the following WordprocessingML:

```
<w:pPr>  
<w:textFlow w:val="btLr" />  
</w:pPr>
```

The textFlow element specifies via the btLr value in the val attribute that the text flow should go bottom to top, and left to right. end example]

The text and examples mention a textFlow element, but the clause is about a textDirection element.

Proposed change: correct text and/or examples and/or schema.

Comment 8

Part 4: throughout

Case mismatch in elements and attribute names are given all throughout Part 4. Examples:

Part 4, Section 2.3.1.33: "...If the beforeAutoSpacing attribute is also specified, then this attribute value is ignored"

In the schema annexed is referenced as "beforeAutospacing"

Part 4, Section 2.3.2.24: "If the csTheme attribute is also specified, then this attribute shall be ignored and that value shall be used instead."

In the schema annexed is referenced as "cstheme"

Part 4, 2.5.2.6 dataBinding (XML Mapping): "The custom XML data identifier, specified using the storeItemID attribute of the dataStoreItem element"

In the schema annexed is referenced as "datastoreItem"

Proposed change: review all the occurrences of elements and attribute names in normative and informative text of Part 4, and assure they match the corresponding schema submitted as DIS 29500's annexes.

Comment 9

Part 4: throughout

Section 2.3.2.8 eastAsianLayout (East Asian Typography Settings) reads:

"...id (East Asian Typography Run ID) Specifies a unique ID which shall be used to link multiple runs containing eastAsianLayout element to each other to ensure that their contents are correctly displayed in the document.

This means that multiple runs which are broken apart due to differences in formatting can be identified as belonging to the same grouping in terms of eastAsianLayout properties, although they are separated into multiple runs of text.

[Example: Consider the following three runs in a document:

```
<w:r>
<w:rPr>
<w:asianLayout w:id="-1552701694" w:combine="lines"
w:combineBrackets="curly" />
</w:rPr>
<w:t>two</w:t>
</w:r>
<w:r>
<w:rPr>
<w:u w:val="single" w:color="4F81BD" w:themeColor="accent1"
/>
<w:asianLayout w:id="-1552701694" w:combine="lines"
w:combineBrackets="curly" />
</w:rPr>
<w:t>lines in</w:t>
</w:r>
<w:r>
<w:rPr>
<w:asianLayout w:id="-1552701694" w:combine="lines"
w:combineBrackets="curly" />
</w:rPr>
<w:t>one</w:t>
</w:r>
```

Although there are three runs of content, all three regions shall be combined into a single two lines in one region based on the identical value used in the id attribute for all three runs. end example]

This is a definition of an attribute of the eastAsianLayout element, but the example references another element: w:asianLayout extraneous to this subclause and the whole Part 4.

This kind of errors are found all throughout Part 4. Examples:

"2.4.73 top (Table Cell Top Margin Exception):

This top cell border is specified using the following WordprocessingML:

```
<w:tc>
<w:tcPr>
...
<w:tcBorders>
<w:top w:val="thinThickThinSmallGap" w:sz="24" w:space="0"
w:color="FF0000"/>
</w:tcBorders>
</w:tcPr>
<w:p/>
</w:tc>
```

The top element specifies a three point border of type thinThinThickSmallGap. end example]"

The example says "thinThickThinSmallGap", but the text says "thinThinThickSmallGap".

"2.3.3.18 noBreakHyphen (Non Breaking Hyphen Character): This element specifies that a non breaking hyphen character shall be placed at the current location in the run content. A non breaking hyphen is the equivalent of Unicode character 002D (the hyphen-minus), however it shall not be used as a valid line breaking character for the current line of text when displaying this WordprocessingML content.

...

If this was not desired, the non breaking hyphen character could be specified as follows:

```
<w:r>
<w:t>This makes a very very very wordy and deliberately overcomplicated
s</w:t>
<w:nonBreakHyphen/>
<w:t>entence.</w:t>
</w:r>
```

This would display a hyphen character, but would not allow the text to break at that location:

This makes a very very very wordy and deliberately overcomplicated s-entence. end example]"

The definition is about the noBreakHyphen element, but the example contain an <w:nonBreakHyphen/> , it is another element, it is a typo?.

"2.18.26 ST_EdnPos (Endnote Positioning Location): This simple type specifies the possible positions of endnotes in a document.

[Example: Consider a document in which endnotes shall be positioned at the end of the section. The section properties for this section shall be declared as follows:

```
<w:settings>
<w:endnotePr>
<w:pos w:val="endSect" />
</w:endnotePr>
...
```

</w:settings>

The val attribute is endSect, therefore the position of endnotes is specified to be at the end the section. end example]

Enumeration Value:

sectEnd (Endnotes Positioned at End of Section)"

The text and XML fragment read "endSect" but the enumeration names it "SectEnd".

Proposed change: Examples of OOXML markup with invalid XML or with typographical errors in elements and attributes names give poor value to the reader and could result in confusion rather than help to understand this specification. It is suggested to review all the examples of Part 4 (aprox. 5500) to catch and correct this kind of errors. Some XML parsing and validating tools are available on the internet that could help in this task (example: saxon, libxml, Microsoft MSXML, etc.), some of them at no cost to the user.

Comment 10

Part 4: throughout

Errors in elements and attribute names are found all throughout **normative** text of Part 4.

Examples:

Part 4, Section 2.9.10 lvlPicBulletId (Picture Numbering Symbol Definition Reference) reads: *"This element specifies a picture which shall be used as a numbering symbol for a given numbering level by referring to a picture numbering symbol definition's numPictBullet element"*

But the numPictBullet element is named "numPicBullet" in the clause 2.9.21

Part 4, Section 2.15.1.18 characterSpacingControl (Character-Level Whitespace Compression) reads: *"The characterSpacingControl element has a val attribute value of 'dontCompress', which specifies that no character compression shall be applied "*

But the dontCompress element is named "doNotCompress" in the schema annexed.

The section 2.15.1.10 names in six contiguous lines of the same page (page 1118) the "Automatically Hyphenate Document Contents When Displayed" element with three different names:

"autoHypehenation" in line 18
"autoHypehnation" in line 20
"autoHyphenation" in line 14

Proposed change: review the entire Part 4 to find and correct all the occurrences of elements and attribute names and and assure they are given an unique name through out Part 4. The name must match the corresponding schema submitted as DIS 29500's annexes.

Comment 11

Part 4, Section 5.1.12.60 ST_TextAnchoringType (Text Anchoring Types) reads:

"This is different than 'anchorJustified'"

No definition of the term 'anchorJustified' was found in Part 4.

Proposed change: define the term, otherwise correct the text.

Comment 12

Part 4, Section 5.1.12.60 ST_TextAnchoringType (Text Anchoring Types) reads:

"When text is horizontal, this spaces out the actual lines of text and is almost always identical in behavior to 'anchorJustified'"

No definition of the term 'anchorJustified' was found in Part 4.

Proposed change: define the term, otherwise correct the text.

Comment 13

Part 4, Section 5.1.12.60 ST_TextAnchoringType (Text Anchoring Types) reads:

"This is different than 'anchorDistributed'"

No definition of the term 'anchorDistributed' found in Part 4.

Proposed change: define the term, otherwise correct the text.

Anexo 4 - Ejemplos ambigüedades normativas/informativas en ECMA-376-OOXML

Comment 1

Part 1, Part 2, Part 3, Part 4, Part 5: Throughout

There seems to be no editorial coherence to mark up text as "notes". In some cases text marked up as notes seems to be normative (even though Part 1 clause 7 says that [note: end note]s are informative). Examples:

"2.7.3.16 semiHidden (Hide Style From Main User Interface)

...

[Note: The interpretation of a "main" user interface shall not be dictated by this Office Open XML Standard, and may be defined by an application as appropriate. This setting is intended to define a style property which allows styles to be seen and modified in an advanced user interface, without exposing the style in a less advanced setting, for example, the style which is used to format the contents of a comment should typically not be shown in a simple user interface (as it is uncommon to want to modify it), but would be inappropriate to hide completely using the hidden element (§2.7.3.4), as very advanced users may want to change its appearance. end note]"

"2.15.1.71 rsidRoot (Original Document revisión Save ID) This element specifies the revisión save ID which was associated with the first editing session for this document.

[Note: This information shall be identical between any number of copies of the same document, as they all originate from the same original editing session. Applications may use this information as desired. end note]"

"3.11.1.21 rrc (revisión Row Column Insert Delete)

...

ref (Reference) A reference to the location of the rows/columns that were inserted or deleted.

[Note: A reference to a whole column or row shall include both the column and row components. For example, column A is referenced by "A1:A1048576", and row 1 is referenced by "A1:XFD1". However, because this attribute value is occurring in the context of an entire row or column insert, the column component of a row reference can be ignored, and the row component of a column reference can be ignored. end note]"

"5.1.2.1 Core Drawing Object Information

...

[Note: Measurement Units - Length units shall be expressed in device-independent physical units: English Metric units (EMUs), points, picas, and inches. Device-dependent units such as pixels shall not be used. end note]"

In other cases, markup is not used to identify notes, so doubt arises about the text being informative or normative.

Examples:

"3.11.1.21 rrc (revisión Row Column Insert Delete)

...

eol (End Of List) A Boolean flag indicating that a row or a column is being inserted at the end of a list of data.

Note: List in this context does not mean table, rather it refers to the feature where the spreadsheet application automatically creates an internal structure for making data input more consistent on

adjacent rows or columns. For instance, if 3 cells in a row are entered with the same format, then when entering data into the 4th adjacent cell, the spreadsheet application might automatically apply that same format. In this case, those cells are treated as a list.

The possible values for this attribute are defined by the XML Schema boolean datatype."

"3.3.1.47 ignoredError (Ignored Error)

...

Note: more than one kind of error can exist on a cell. These flags are not mutually exclusive."

"3.10.1.3 cacheField (PivotCache Field)

...

numFmtId (Number Format Id) Specifies the number format that is applied to all items in the field.

Number formats are written to the styles part. For more information see §3.8.31 in Styles.

Note: Formatting information provided by cell table and by PivotTable need not agree. If the two formats differ, the cell-level formatting takes precedence. If you change the layout of the PivotTable, the PivotTable formatting will then take precedence.

The possible values for this attribute are defined by the ST_NumFmtId simple type (§3.18.49)."

The same goes to the examples markup: identical text is marked up as example in a subclause but in other is leaved without mark up. For example:

"2.7.3 General Style Properties:

General style properties refer to the set of properties which can be used regardless of the type of style.

[Example: Within a style definition the style name, additional aliases

for the style, a style ID (used by the document content to refer to the style), if style is hidden, if style is locked, etc. are general style properties. end example]"

(text marked up as example)

"2.7.3.17 style (Style Definition)

...General style properties refers to the set of properties which can be used regardless of the type of style; for example, the style name, additional aliases for the style, a style ID

(used by the document content to refer to the style), if style is hidden, if style is locked, etc "

(the same text without example mark up, using a "for example" wording)

The lack of editorial uniformity can result in confusion to the reader of this specification and raise doubt about informative and normative text.

Proposed change:

i) Adopt an unique and coherent criteria to mark text as notes and stick to it in all DIS 29500.

ii) Adopt an unique and coherent criteria to mark text as examples and stick to it in all DIS 29500.

Comment 2

Part 3: throughout

Overlapping "start" and "end" marks of informative text were found in Part 3.

Examples:

Page 229 (4. Introduction to PresentationML) contains a "This clause is informative" begin mark, but the immediate prior mark (at page 92) was a "This clause is informative" begin mark too.

Page 260 (5. Introduction to DrawingML) contains a "End of informative text." end mark, but the immediate prior mark (at page 258) was a "End of informative text." end mark too.

So, there are big portions of Part 3 where there is no way to tell if they are "normative" or "informative" text.

Proposed change: revise all the text of Part 3 and correct the overlapping informative marks.

Comment 3

Part 1, Section 7 is a normative clause, and reads:

"General Description

...

Clauses 1–5, 7, and 9–14 form a normative part of this Part; and the Introduction, clauses 6 and 8, as well as the annexes, notes, examples, rationale, guidance, and the index, are informative."

But, conform the "This clause is informative" and "End of informative text" marks, the clause 15 ("Shared") also should be a normative part, and it is not listed in the phrase "Clauses 1–5, 7, and 9–14 form a normative part of this Part"

Proposed change: clarify if clause 15 normative or informative text.

Comment 4

Part 1, Section 7, which is a normative clause, reads:

"General Description

...

This Part is divided into the following subdivisions:

- 1. Front matter (clauses 1–7);*
- 2. Overview (clause 8);*
- 3. Main body (clauses 9–14);*
- 4. Annexes*

...

Clauses 1–5, 7, and 9–14 form a normative part of this Part; and the Introduction, clauses 6 and 8, as well as the annexes, notes, examples, rationale, guidance, and the index, are informative."

Note that the clause 15 (page 136) is not mentioned in both the list of subdivisions nor the normative and informative clause qualification.

Proposed change: clarify if clause 15 is part of DIS 29500 - Part 1. If it is, then harmonize the text of Part 1 Section 7 to reflect that.

Comment 5

Part 2, Section 6, which is a normative clause, reads:

"General Description

This Open Packaging specification is intended for use by implementers, academics, and application programmers. As such, it contains a considerable amount of explanatory material that, strictly speaking, is not necessary in a formal specification.

This Open Packaging specification is divided into the following subdivisions:

- 1. Front matter (clauses 1–7);*
- 2. Overview (clause 8);*
- 3. Main body (clauses 9-13);*
- 4. Annexes"*

A clause 13 is mentioned but it wasn't found in Part 2.

Proposed change: provide the missing clause 13, or if there isn't one, then harmonize the text of Part 2, Section 6 to cite the clauses actually included.

Comment 6

Part 2, Section 6, which is a normative clause, reads:

"General Description

This Open Packaging specification is intended for use by implementers, academics, and application programmers. As such, it contains a considerable amount of explanatory material that, strictly speaking, is not necessary in a formal specification.

This Open Packaging specification is divided into the following subdivisions:

- 1. Front matter (clauses 1–7);*
- 2. Overview (clause 8);*
- 3. Main body (clauses 9-13);*
- 4. Annexes"*

This text says that clause 8 is "Overview", but the clause 8 (page 11) says "Package model" and the "Overview" clause has the number 7.

Proposed change: harmonize the text of Part 2 with the correct title and content of subclauses.

Comment 7

Part 5, Section 6, which is a normative clause, reads:

"6. General Description

This specification is intended for use by implementers, academics, and application programmers. As such, it contains a considerable amount of explanatory material that, strictly speaking, is not necessary in a formal specification.

This specification is divided into the following subdivisions:

1. *Front matter (clauses 1–7);*
2. *Overview and introductory material (clause 8);*
3. *Main body (clauses 9–13);*
4. *Annexes"*

This text says that clause 8 is "Overview and introductory material", but the clause 8 (page 9) title is "8. Markup Compatibility Fundamentals" and the "Overview" clause has the number 7.

Also, a clause 13 is mentioned but it wasn't found in Part 5.

Proposed change: harmonize the text of Part 5 with the correct title and content of clauses. Provide the clause 13, if it is missing.

Comment 8

Part 4, Section 5.3 DrawingML - Legacy Compatibility

The normative text of the complete Section 5.3 reads:

"Within the context of DrawingML, it must be possible (for considerations to legacy compatibility) to be able to include explicit references to specific shapes within VML Drawing parts.

5.3.2 Basics

Legacy Compatibility is part of the shape definitions and properties of the DrawingML framework.

5.3.2.1 1 legacyDrawing (Legacy Drawing Object)

This element specifies the shape ID for a legacy drawing object. These legacy drawing objects all have a shape ID associated with them that is unique across the entire document. In order to store these legacy shape IDs as well as new shape IDs this legacyDrawing element should be used.

Attributes: spid (Shape ID): Legacy Shape ID that is unique throughout the entire document. Legacy shape IDs should be assigned based on which portion of the document the drawing resides on. The assignment of these ids is broken down into clusters of 1024 values. The first cluster is 1-1024, the second 1025-2048 and so on."

There are two problems with this text:

i) the first paragraph says "This element specifies the shape ID for a legacy drawing object." but the same paragraph says later "In order to store these legacy shape IDs as well as new shape IDs this legacyDrawing element should be used.". So it is no clear if this element specifies shape IDs of legacy drawing objects only, or for legacy objects and new (no legacy) objects.

ii) The 2nd paragraph says "Legacy shape IDs should be assigned based on which portion of the document the drawing resides on" but gives no indication of how to perform this assignment. There are

three examples that mention one criteria of assignment but this examples are informative (besides that, the examples provided later contain the word "should", so it raise the doubt if the examples are indeed normative and have incorrectly been marked as examples).

Proposed change: revise all the text of 5.3.2 to clarify the definition of the "Legacy Drawing Object", and provide precise normative of how to assign shape IDs to them.

Fernando Adrián Toledo
Presidente
DNI 26.010.013
Asociación Civil SOLAR
Software Libre Argentina