ODF Advocacy Open Project

OOXML Non Technical and Technical Issues

Italo Vignoli

OOXML Transitional and Strict

- As of 2020, the Office default for .docx, .xlsx and .pptx is Transitional OOXML, a proprietary document format which was created as a bridge from legacy MS Office formats and the approved ISO Standard.
- OOXML Strict is the ISO approved open standard, but being the non publicized last option available from MS Office “file, save as…” menu has never been adopted, so 100% of existing OOXML files we are referring to are proprietary.

OOXML Strict Standard Support

- MS Office 2010: NO
- MS Office 2013: YES, but default is Transitional
- MS Office 2016: YES, but default is Transitional
- MS Office 2019: YES, but default is Transitional
- MS Office macOS: NO
- MS Office 365: NO
- According to Microsoft statements in 2007, OOXML Strict should have been the default since Office 2010

De Jure vs De Facto Standards

- A **de facto** standard refers to a significant market share
- A **de jure** standard is based on a collective agreement
- As such they are innately different, as are their value and effect on the market
- **De jure standards for document formats**
  - Foster interoperability, create network externalities, prevent lock-in, cut transaction costs, create a transparent market and reduce variety
- **De facto standards for document formats**
  - Tend to be the exact opposite, to increase supplier-dependence and create an obfuscated market

One vs Two De Jure Standards

	Effect on the Market	
	One Standard	**Two Standards**
Information	• Increases market transparency • Reduces transaction costs (and information asymmetry) • Correct wrong selection of tool • Facilitates exchange	• Reduce market transparency • Increase transaction costs (need of converters and converting) • Make comparison of product quality more difficult • Hinder trade with higher information transaction costs
Compatibility	• Creates network externalities and increases competition • Decreases vendor lock-in	• Reduce interoperability and adds switching costs • Reduce network externalities and decrease competition • Increase likelihood of standard based vendor lock-in
Variety Reduction	• Allows economies of scale • Facilitates building a critical mass	• Reduced economies of scale • Reduced chanches of building a critical mass

ODF Philosophy

- The philosophy behind the ODF standard document format was to design a mechanism in a vendor neutral manner from the ground up using existing standards wherever possible
- Although this means that software vendors would need to tweak their individual packages more than if they continued down their original routes the benefits for interoperability were important enough to justify the move

OOXML Philosophy

- The OOXML pseudo-standard document format appears to be designed by Microsoft for Microsoft products, and to inter-operate with the Microsoft environment
- Little thought appears to have been exercised for interoperability with non-Microsoft environments or compliance with established vendor-neutral standards

ODF vs OOXML Strategic Difference

- ODF has been designed as a document standard for the next 20-50 years, to liberate users from the lock-in strategy built into yesterday's and today's proprietary formats, and foster interoperability
- OOXML has been designed as a pseudo-standard document format to propagate yesterday's document issues and lock-in strategy for the next 20-50 years, to the detriment of users and interoperability

Standardization Process

ODF

- Based on OOo XML format
- Dec 12, 2002: document format presented to OASIS
- May 1, 2005: ODF released by OASIS
- Nov 16, 2005: ODF presented to ISO/IEC JTC1 based on Publicly Available Specification (PAS)
- May 3, 2006: ODF approved as ISO/IEC IS 26300 standard
- Review: 720 paged in 1239 days

OOXML

- Based on Microsoft Office 2003 XML format
- Dec 15, 2005: document format presented to ECMA
- Dec 31, 2006: ECMA standard approved by General Assembly
- Jan 31, 2007: OOXML presented to ISO/IEC JTC1 based on Fast Track
- Mar 31, 2008: OOXML pseudo standard approved
- Review: 7200 pages in 838 days

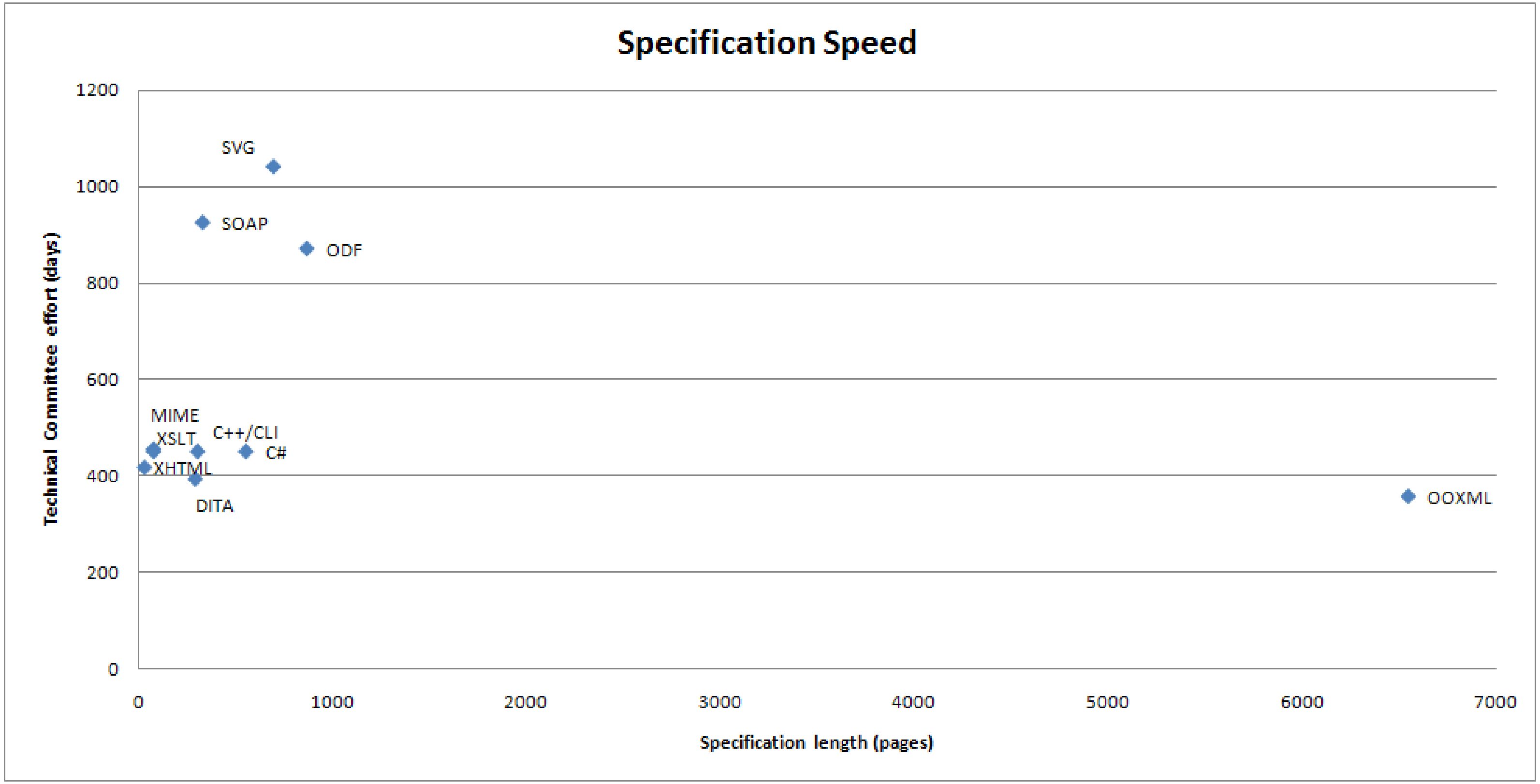
OOXML Approval at Record Speed

- 6000 pages reviewed in 30 days @200pages/day
- ECMA 376 was prepared in a hurry, with a calculated page review/edit/approve rate approximately 20 times faster than other markup standards
- The time available to review the gigantic specification was not sufficient
- ECMA 376 was finalized by ECMA on December 7 and submitted to JTC-1 less than 30 days later

ISO Rule for Specification Review

- ISO/IEC JTC 1 Directives, Edition 5, Version 2.0 states that in relation to PAS submissions: "The specification shall have had sufficient review over an extended time period to characterise it as being stable" (JTC1 Directives, Annex M, The Transposition of Publicly Available Specifications into International Standards, A Management Guide, M.7.4.1.3)
- Since the specification was submitted for fast-track resolution almost immediately after its development, and its development was behind closed doors, this requirement has not been met

Specification Speed Comparison



Reuse of Existing Standards

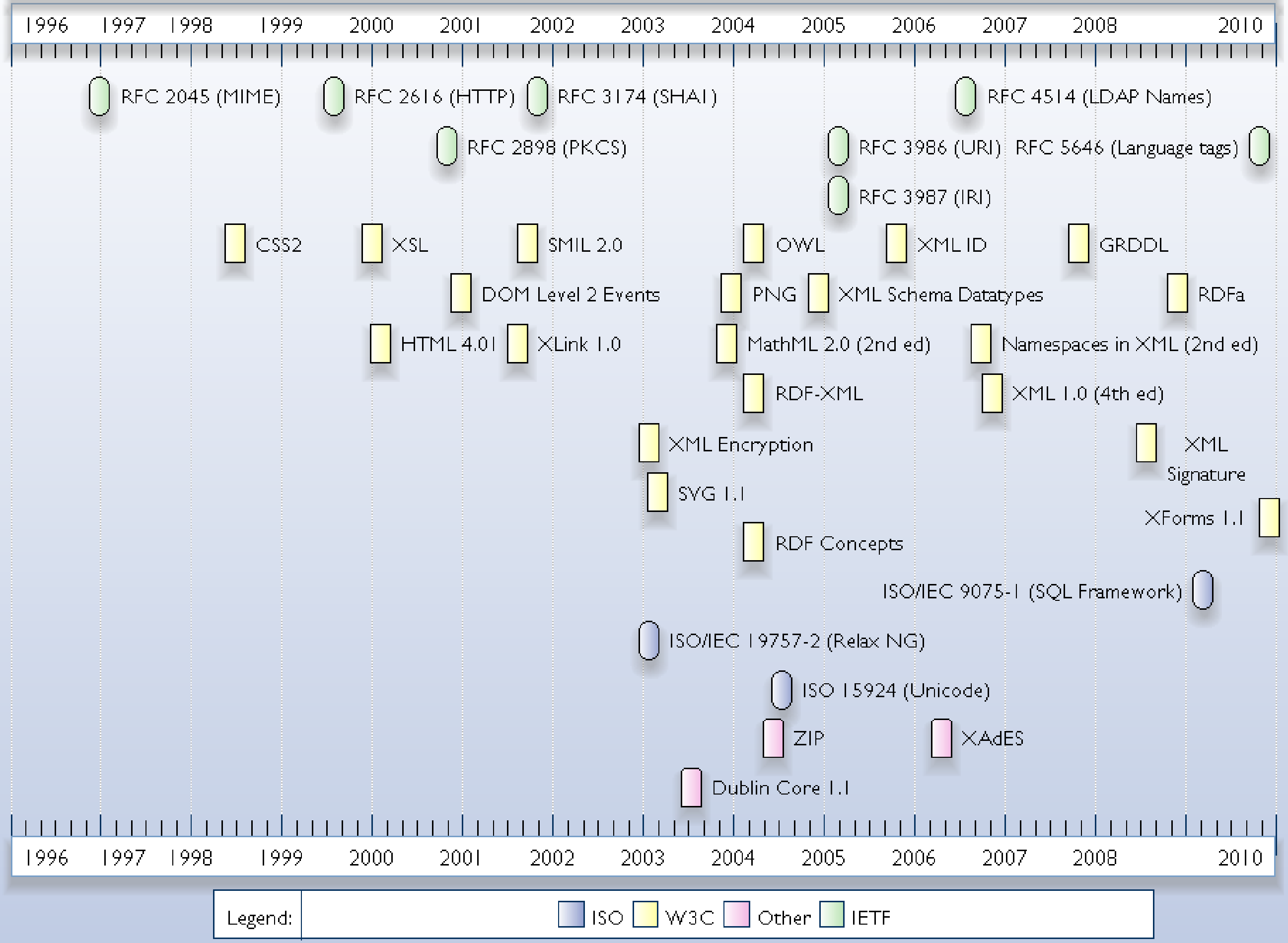
ODF

- Dublin Core
- XLS:FO
- SVG
- MathML
- XLink
- SMIL
- XForms

OOXML

- Dublin Core

ODF Normative References



1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2010

RFC 2045 (MIME)

RFC 2616 (HTTP)

RFC 3174 (SHA1)

RFC 4514 (LDAP Names)

RFC 2898 (PKCS)

RFC 3986 (URI)

RFC 5646 (Language tags)

RFC 3987 (IRI)

CSS2

XSL

SMIL 2.0

OWL

XML ID

GRDDL

DOM Level 2 Events

PNG

XML Schema Datatypes

RDFa

HTML 4.01

XLink 1.0

MathML 2.0 (2nd ed)

Namespaces in XML (2nd ed)

RDF-XML

XML 1.0 (4th ed)

XML Encryption

XML Signature

SVG 1.1

XForms 1.1

RDF Concepts

ISO/IEC 9075-1 (SQL Framework)

ISO/IEC 19757-2 (Relax NG)

ISO 15924 (Unicode)

ZIP

XAdES

Dublin Core 1.1

1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2010

Legend: 🟦 ISO 🟨 W3C 🟪 Other 🟩 IETF

LibreOffice as Shakespeare (ODT)

2017

```
<text:p text:style-name="P1">To be, or not to be,
that is the question</text:p>
```

2018

```
<text:p text:style-name="P1">To be, or not to be,
that is the question</text:p>
```

2019

```
<text:p text:style-name="P1">To be, or not to be,
that is the question</text:p>
```

MS Office as Shakespeare (DOCX)

2017

```
<w:t>To be</w:t>
<w:t>,</w:t>
<w:t xml:space="preserve"> or not to be</w:t>
<w:t>,</w:t>
<w:t xml:space="preserve"> that </w:t>
<w:t>is the question</w:t>
```

MS Office as Shakespeare (DOCX)

2018

```
<w:t>To be</w:t>
<w:t>, or</w:t>
<w:t xml:space="preserve"> not to be</w:t>
<w:t>,</w:t>
<w:t xml:space="preserve"> that is the</w:t>
<w:t xml:space="preserve"> question</w:t>
```

MS Office as Shakespeare (DOCX)

2019

```
<w:t>To be</w:t>
<w:t>,</w:t>
<w:t xml:space="preserve"> or not to be, that </w:t>
<w:t>is the question</w:t>
```

MS Office as Shakespeare (DOCX)

2020 (Office 2019)

```
<w:t>To be, or not to b</w:t>
<w:t>e, that is the question.</w:t>
```

MS Office as Shakespeare (DOCX)

2020 (Office 365)

```
<w:t xml:space="preserve">To be, or </w:t>
<w:t>not</w:t>
<w:t xml:space="preserve"> to be, </w:t>
<w:t>that</w:t>
<w:t xml:space="preserve"> </w:t>
<w:t>is</w:t>
<w:t xml:space="preserve"> the question.</w:t>
```

Brain & Computer

Brain
Red

Computer
#FF0000

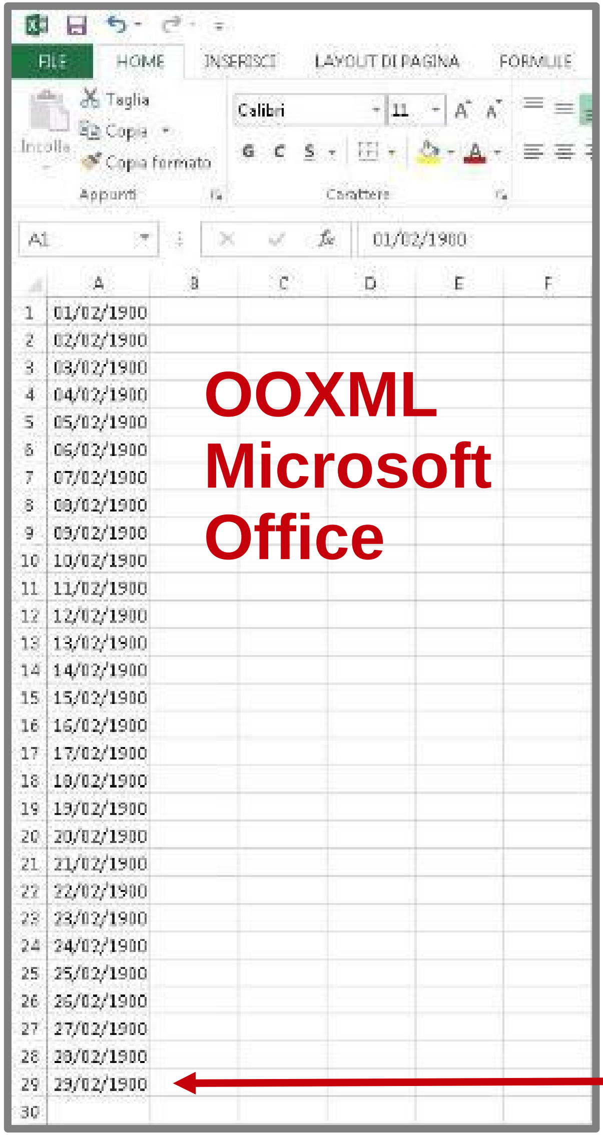
Brain & Computer

ODF (LibreOffice)

- Writer
  fo:color="#FF0000"
- Calc
  fo:color="#FF0000"
- Impress
  fo:color="#FF0000"

OOXML (MS Office)

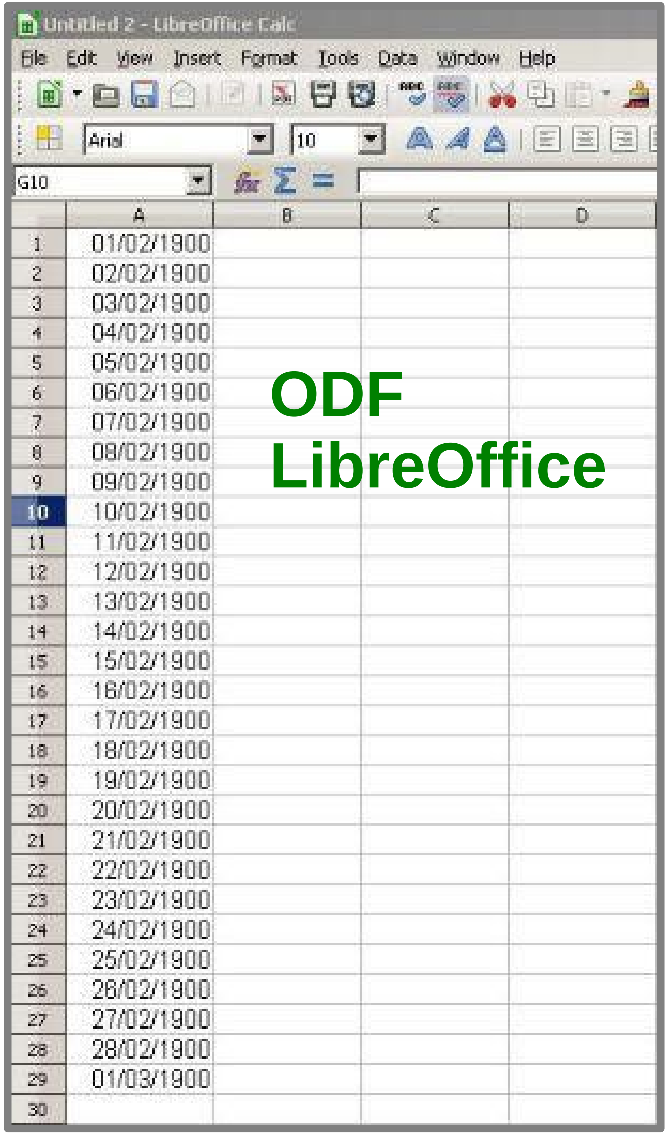
- Word
  w:color w:val="FF0000"
- Excel
  color rgb="FFFF0000"
- PowerPoint
  a:srgbClr val="FF0000"



A1 | 01/02/1900

	A
1	01/02/1900
2	02/02/1900
3	03/02/1900
4	04/02/1900
5	05/02/1900
6	06/02/1900
7	07/02/1900
8	08/02/1900
9	09/02/1900
10	10/02/1900
11	11/02/1900
12	12/02/1900
13	13/02/1900
14	14/02/1900
15	15/02/1900
16	16/02/1900
17	17/02/1900
18	18/02/1900
19	19/02/1900
20	20/02/1900
21	21/02/1900
22	22/02/1900
23	23/02/1900
24	24/02/1900
25	25/02/1900
26	26/02/1900
27	27/02/1900
28	28/02/1900
29	29/02/1900
30	

OOXML
Microsoft Office



G10

	A
1	01/02/1900
2	02/02/1900
3	03/02/1900
4	04/02/1900
5	05/02/1900
6	06/02/1900
7	07/02/1900
8	08/02/1900
9	09/02/1900
10	10/02/1900
11	11/02/1900
12	12/02/1900
13	13/02/1900
14	14/02/1900
15	15/02/1900
16	16/02/1900
17	17/02/1900
18	18/02/1900
19	19/02/1900
20	20/02/1900
21	21/02/1900
22	22/02/1900
23	23/02/1900
24	24/02/1900
25	25/02/1900
26	26/02/1900
27	27/02/1900
28	28/02/1900
29	01/03/1900
30	

ODF
LibreOffice

Writing Dates the Excel Way

Event	Calc	Excel
Italo Vignoli Birthday	12/08/1954	19948
Italo Vignoli Graduation	19/11/1978	28813
Italo Vignoli First Job	01/10/1981	29860
Italo Vignoli First Computer	01/09/1983	30560
Italo Vignoli Wedding	08/09/1984	30933
Italo Vignoli Installs OOo	02/01/2003	37623
LibreOffice Announcement	28/09/2010	40449

Invalid Date Calculations

- Treats 1900 as leap year, an old bug in Excel (as per MS: inherited from Lotus 1-2-3)
- This contradicts the Gregorian calendar, ISO 8601 and the civil calendar adopted by most nations of the world
- Gregorian Calendar says: Years divisible by 100 are leap years only if they are also divisible by 400, which 1900 is not, clearly
- Unacceptable to propagate ancient bugs into a forward looking proposed standard

OOXML Contradicts ISO 639

- ISO 639 is the set of ISO standards that lists short codes for language names
- OOXML uses its own fixed list of numbers (ECMA 376 section 2.18.52, page 2530, ST_LangCode)
- Data interchange is affected when communicating with non-MS software which are written to adhere to standards

OOXML Conflicts ISO 8632

- ISO/IEC 8632 is the ISO standard for computer graphics metafiles: "2D graphical (pictorial) information" consisting of "vector graphics", "raster graphics", and "text" (NIST, 1998)
- OOXML recommends Windows Metafiles or Enhanced Metafiles instead of using ISO/IEC 8632 or W3C SVG
- WMF are Windows-only proprietary formats, and they have not been approved as ISO standards
- Why use a proprietary standard when an International standard exists?

OOXML Conflicts with W3C SVG

- W3C SVG is the W3C standard "for describing two-dimensional vector and mixed vector/raster graphics in XML"
- Ecma 376, section 14 (page 132) "DrawingML", defines a vector drawing XML format in conflict with the industry standard W3C SVG
- Ecma 376, section 8.6.2 (page 24) "VML", requires support for another drawing XML format in conflict with W3C SVG
- VML was proposed by Microsoft as a W3C standard in 1998, but was rejected in favour of SVG and deprecated by Microsoft

OOXML Poor XML

- Poor names and inconsistent naming conventions for elements and attributes
- Ecma 376 contradicts the goals of XML which are
  - XML documents should be human-legible and reasonably clear
  - Terseness in XML markup is of minimal importance
- Instead, Ecma 376 often uses unclear names and inconsistent naming conventions
  - These include unnecessary vowel removals, name truncations, and unusual abbreviations, as described in next slide

Examples of Poor XML

- In VML (5.1.10.45, page 4413) "outerShdw" (Outer Shadow Effect) has its second word devoid of vowels, but its Child Elements and Attributes have different naming conventions: e.g. scrgbClr, algn, blurRad, dir, dist, rotWithShape
- In WordprocessingML (2.15.1.78, page 2020) "settings" (Document Settings) has a large list of Child Elements, and within that it has significant contradictory naming conventions, e.g. ActiveWritingStyle, attachedSchema, documentType, docVars, endnotePr, hdrShapeDefaults

OOXML Plain Text Flaws

- Unspecified terms exist for “plain text”
- Ecma 376 section 11.3.1 (page 38) "Alternative Format Import Part", allows content in "plain text"
- The encoding for "plain text" is not specified (is it 7-bit ASCII? ISO 8859-1? UTF-8?), and as such it will not allow international interoperable use
- This is serious problem since XML document standards may be used by non-US -English implementations

OOXML Conflicts W3C MathML

- MathML is the W3C standard for "describing mathematical notation and capturing both its structure and content"
- Ecma 376, section 7.1 (page 747) "Math", covers mathematical expressions, and defines a format in conflict and incompatible with the W3C recommended MathML
- MathML is included in the ISO/IEC 26300 standard (ODF) in section 12.5 "Mathematical Content", and as a result Ecma 376 conflicts with an ISO specification for mathematical notation

OOXML Hash and Encryption

- OOXML ignores ISO/IEC 10118-3, W3C XML-ENC, and other cryptographic hash standards
- Ecma 376 ignores accepted standards for cryptographic hashes and defies expert standards for cryptography, by proposing its own hash algorithms which are almost certainly flawed
- Cryptography is a hard subject, algorithms & implementations need to go through expert- and peer-review to be considered safe for use
- See what Bruce Schneier, well known security expert has to say: http://www.schneier.com/crypto-gram-9904.html

Hash and Encryption Background

- ISO has chosen the "Whirlpool" algorithm as standard ISO 10118-3
- The W3C, in its XML-ENC standard, includes a list of algorithms: SHA1, SHA256, SHA512, RIPEMD-160
- The European NESSIE project recommends: ISO 10118-3 ("Whirlpool"), SHA-256, SHA-384 and SHA-512
- In the USA, NIST recommends SHA1, SHA224, SHA256, SHA384, and SHA512
- In Japan, CRYPTREC recommends MD5, RIPEMD-160, SHA1, SHA256, SHA384, and SHA512

OOXML and Encryption

- Ecma 376, section 2.15.1.28 (page 1941), does not follow the advice of any of these organizations, and defines new hashing algorithms that have not undergone scrutiny by the cryptographic community
- Section 2.15.1.28 (page 1941) defines one, while Sections 3.3.1.69 (page 2786) "protectedRange" and 3.2.29 (page 2698) define another very similar algorithm

OOXML Conflicts W3C SMIL

- Synchronized Multimedia Integration Language (SMIL) is the W3C standard for "synchronized multimedia presentation"
- With SMIL an author can:
  - Describe the temporal behavior of the presentation
  - Describe the layout of the presentation on a screen
  - Associate hyperlinks with media objects
- Ecma 376, section 4.4 (page 565) "Animation", covers presentation animations (slide transitions), in conflict with the W3C Recommendation SMIL

Proprietary Unit of Measurement

- ECMA 376 Fabricates units of measurement
- Attributes throughout the ECMA 376 specifications take values in "English Metric Units" (EMU): for example, attributes of type ST_PositiveCoordinate (5.1.12.42, page 4505) are measured in EMUs
- This is not a known unit in existing literature, and is only defined inside a paragraph in section 5.9.2.1 (page 655), so that "91440 EMUs/US inch or 36000 EMUs/cm"
- Similarly, section 2.18.105 (page 1836), specifies "twips" or twentieths of a point (1/1440th of an inch)

OOXML Internal Inconsistencies

- The w:sz element is an example of internal inconsistencies in the specifications measurements:
  - For fonts, the w:sz element specifies the size in half points (2.3.2.36, page 1013)
  - For frameset, the w:sz element has a string value that could be a relative value, a percentage, or a number of pixels (2.15.2.39, page 2136), while the examples on page 2138 do not refer to w:sz at all
- However, as the child of rPr (3.4.11, page 2846), its value is in points

Internal Inconsistencies: ST_Border

- Section 2.18.4 (page 2414) lists numerous styles such as apples, scaredCat, heebieJeebies, etc.
- However, the specification does not fully define these styles (e.g missing height, width, color-depth, orientation)
- The style basicThinLine describes behavior for horizontal, vertical and corner scenarios but many styles (e.g babyRattle, balloonsHotair, etc) provide no such details
- The problem with this is that a single style can be interpreted differently by different vendors/implementors, and do not provide generality

OOXML Messes Up Hex Numbers

- Confusing and inconsistent definitions of lengths of hexadecimal numbers
- Ecma 376 uses confusing and inconsistent definitions of values with hexadecimal numbers
- For example, section 2.18.52 (page 2531) ST_LangCode, is defined on the text as a "two digit hexadecimal code", but the values given cannot be represented by only two hexadecimal digits, but needs four

Inflexible Numbering Format

- Section 2.18.66 page 2554, ST_NumberFormat
- Numbering Format for number lists (2.9.18 page 1581), footnotes (2.11.17 page 1645), endnotes (2.11.18 page 1646), captions (2.15.1.16 page 1912), page numbers (2.6.12 page 1412)
- Contradicts W3C XSLT used by ISO 26300
- Contradicts Unicode ISO 10646

Problems with Percentages

- Section 2.18.85 (p. 2583) uses predefined symbols (like "pct15" for 15%) in 5 or 2.5 percent increments (which is inflexible and difficult to process with standard XML tools, compared to a generic number-valued field)
- Section 2.15.1.95 (p. 2053) uses a decimal number giving the percentageSection 2.18.97 (p. 2608) uses a number in 50ths of a percent
- Section 5.1.12.41 (p. 4505) uses a number in 1000ths of a percent

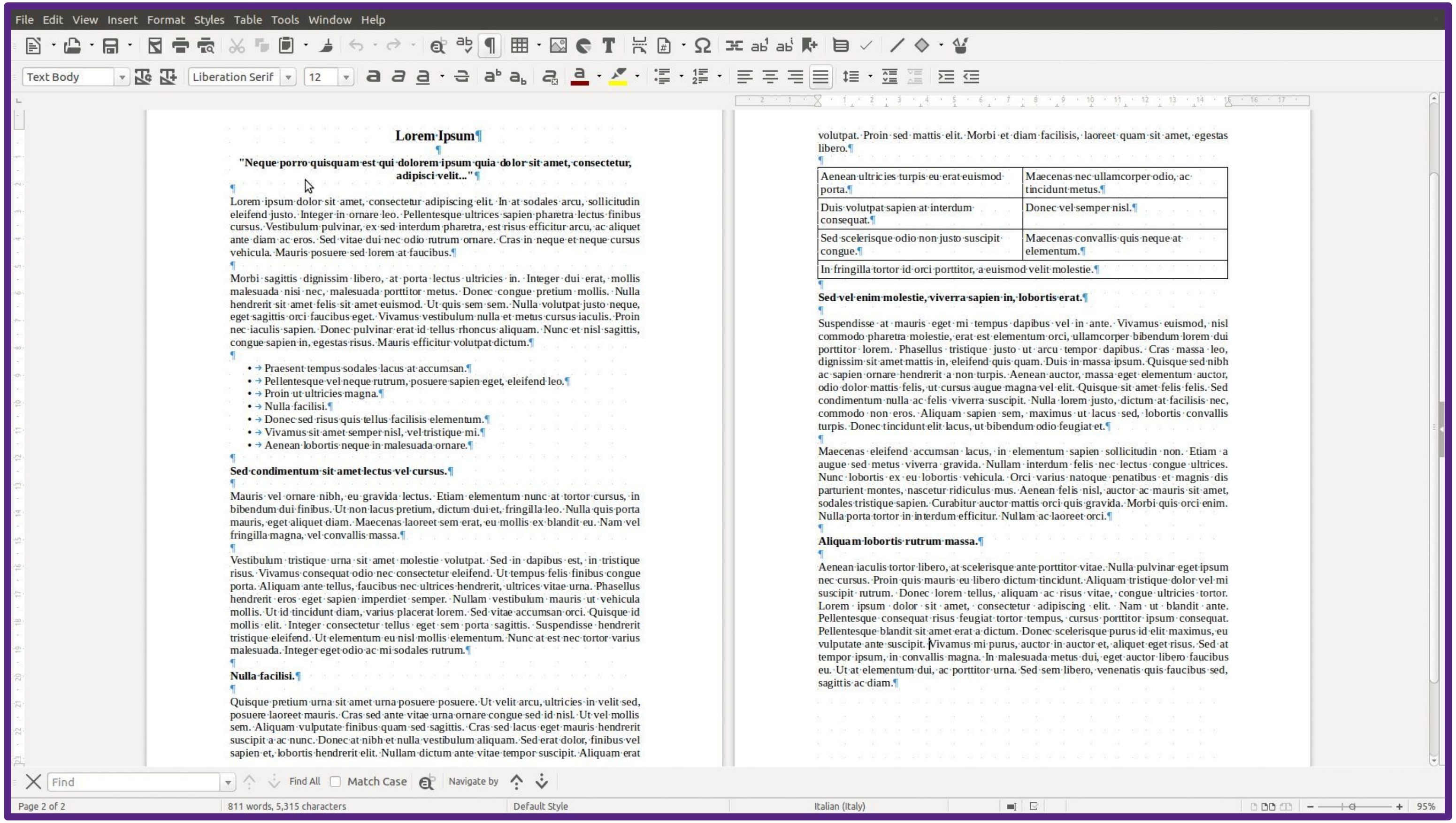
Proprietary Contents & Behaviours

- Section 6.2.3.17 "Embedded Object Alternate Image Requests Types" (page 5679) requires implementors to support the proprietary Windows Metafiles
- Several sections require the implementor to clone the non specified behaviour of a proprietary product:
  - Section 2.15.3.6 page 2161, autoSpaceLikeWord95
  - Section 2.15.3.31 page 2209, lineWrapLikeWord6
  - Section 2.15.3.32 page 2210, mwSmallCaps
  - Section 2.15.3.41 page 2225, shapeLayoutLikeWW8

Redefinition of Colours

- Ecma 376 section 2.18.46 (page 2521) contradicts the standard SVG Color Keyword Names's hexadecimal RGB values for given color names
  - Dark blue: SVG 00008B, Ecma 000080
  - Dark cyan: SVG 008B8B, Ecma 008080
  - Dark gray: SVG A9A9A9, Ecma 808080
  - Dark green: SVG 006400, Ecma 008000
  - Dark red: SVG 8B0000, Ecma 800000
  - Light gray: SVG D3D3D3, Ecma C0C0C0

Comparison ODF/OOXML



Lorem Ipsum

"Neque porro quisquam est qui dolorem ipsum quia dolor sit amet, consectetur, adipisci velit..."

Lorem ipsum dolor sit amet, consectetur adipiscing elit. In at sodales arcu, sollicitudin eleifend justo. Integer in ornare leo. Pellentesque ultrices sapien pharetra lectus finibus cursus. Vestibulum pulvinar, ex sed interdum pharetra, est risus efficitur arcu, ac aliquet ante diam ac eros. Sed vitae dui nec odio rutrum ornare. Cras in neque et neque cursus vehicula. Mauris posuere sed lorem at faucibus.

Morbi sagittis dignissim libero, at porta lectus ultricies in. Integer dui erat, mollis malesuada nisi nec, malesuada porttitor metus. Donec congue pretium mollis. Nulla hendrerit sit amet felis sit amet euismod. Ut quis sem sem. Nulla volutpat justo neque, eget sagittis orci faucibus eget. Vivamus vestibulum nulla et metus cursus iaculis. Proin nec iaculis sapien. Donec pulvinar erat id tellus rhoncus aliquam. Nunc et nisl sagittis, congue sapien in, egestas risus. Mauris efficitur volutpat dictum.

- Praesent tempus sodales lacus at accumsan.
- Pellentesque vel neque rutrum, posuere sapien eget, eleifend leo.
- Proin ut ultricies magna.
- Nulla facilisi.
- Donec sed risus quis tellus facilisis elementum.
- Vivamus sit amet semper nisl, vel tristique mi.
- Aenean lobortis neque in malesuada ornare.

Sed condimentum sit amet lectus vel cursus.

Mauris vel ornare nibh, eu gravida lectus. Etiam elementum nunc at tortor cursus, in bibendum dui finibus. Ut non lacus pretium, dictum dui et, fringilla leo. Nulla quis porta mauris, eget aliquet diam. Maecenas laoreet sem erat, eu mollis ex blandit eu. Nam vel fringilla magna, vel convallis massa.

Vestibulum tristique urna sit amet molestie volutpat. Sed in dapibus est, in tristique risus. Vivamus consequat odio nec consectetur eleifend. Ut tempus felis finibus congue porta. Aliquam ante tellus, faucibus nec ultrices hendrerit, ultrices vitae urna. Phasellus hendrerit eros eget sapien imperdiet semper. Nullam vestibulum mauris ut vehicula mollis. Ut id tincidunt diam, varius placerat lorem. Sed vitae accumsan orci. Quisque id mollis elit. Integer consectetur tellus eget sem porta sagittis. Suspendisse hendrerit tristique eleifend. Ut elementum eu nisl mollis elementum. Nunc at est nec tortor varius malesuada. Integer eget odio ac mi sodales rutrum.

Nulla facilisi.

Quisque pretium urna sit amet urna posuere posuere. Ut velit arcu, ultricies in velit sed, posuere laoreet mauris. Cras sed ante vitae urna ornare congue sed id nisl. Ut vel mollis sem. Aliquam vulputate finibus quam sed sagittis. Cras sed lacus eget mauris hendrerit suscipit a ac nunc. Donec at nibh et nulla vestibulum aliquam. Sed erat dolor, finibus vel sapien et, lobortis hendrerit elit. Nullam dictum ante vitae tempor suscipit. Aliquam erat volutpat. Proin sed mattis elit. Morbi et diam facilisis, laoreet quam sit amet, egestas libero.

Aenean ultricies turpis eu erat euismod porta.	Maecenas nec ullamcorper odio, ac tincidunt metus.
Duis volutpat sapien at interdum consequat.	Donec vel semper nisl.
Sed scelerisque odio non justo suscipit congue.	Maecenas convallis quis neque at elementum.
In fringilla tortor id orci porttitor, a euismod velit molestie.	

Sed vel enim molestie, viverra sapien in, lobortis erat.

Suspendisse at mauris eget mi tempus dapibus vel in ante. Vivamus euismod, nisl commodo pharetra molestie, erat est elementum orci, ullamcorper bibendum lorem dui porttitor lorem. Phasellus tristique justo ut arcu tempor dapibus. Cras massa leo, dignissim sit amet mattis in, eleifend quis quam. Duis in massa ipsum. Quisque sed nibh ac sapien ornare hendrerit a non turpis. Aenean auctor, massa eget elementum auctor, odio dolor mattis felis, ut cursus augue magna vel elit. Quisque sit amet felis felis. Sed condimentum nulla ac felis viverra suscipit. Nulla lorem justo, dictum at facilisis nec, commodo non eros. Aliquam sapien sem, maximus ut lacus sed, lobortis convallis turpis. Donec tincidunt elit lacus, ut bibendum odio feugiat et.

Maecenas eleifend accumsan lacus, in elementum sapien sollicitudin non. Etiam a augue sed metus viverra gravida. Nullam interdum felis nec lectus congue ultrices. Nunc lobortis ex eu lobortis vehicula. Orci varius natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aenean felis nisl, auctor ac mauris sit amet, sodales tristique sapien. Curabitur auctor mattis orci quis gravida. Morbi quis orci enim. Nulla porta tortor in interdum efficitur. Nullam ac laoreet orci.

Aliquam lobortis rutrum massa.

Aenean iaculis tortor libero, at scelerisque ante porttitor vitae. Nulla pulvinar eget ipsum nec cursus. Proin quis mauris eu libero dictum tincidunt. Aliquam tristique dolor vel mi suscipit rutrum. Donec lorem tellus, aliquam ac risus vitae, congue ultricies tortor. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam ut blandit ante. Pellentesque consequat risus feugiat tortor tempus, cursus porttitor ipsum consequat. Pellentesque blandit sit amet erat a dictum. Donec scelerisque purus id elit maximus, eu vulputate ante suscipit. Vivamus mi purus, auctor in auctor et, aliquet eget risus. Sed at tempor ipsum, in convallis magna. In malesuada metus dui, eget auctor libero faucibus eu. Ut at elementum dui, ac porttitor urna. Sed sem libero, venenatis quis faucibus sed, sagittis ac diam.

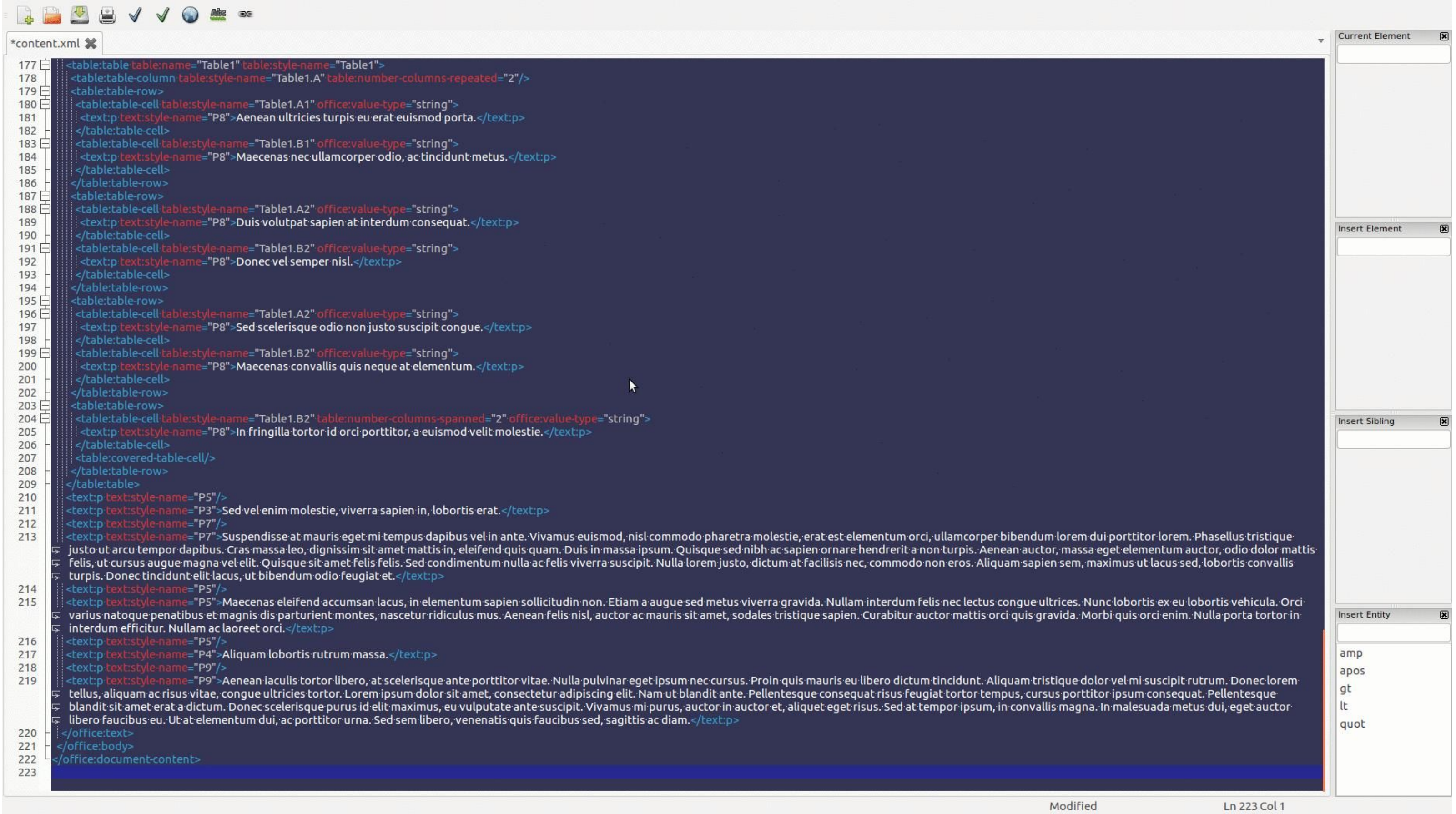
Length in Lines of the XML File

Version	Lines of XML
ODF 1.2 (any version of) LibreOffice	222
OOXML 2010 Transitional (MS Office Windows)	1040
OOXML 2011 Transitional (MS Office MacOS)	12854
OOXML 2013 Transitional (MS Office Windows)	1590
OOXML 2016 Transitional (MS Office Windows)	11667
OOXML 2016 Transitional (MS Office MacOS)	11646
OOXML 2019 Transitional (MS Office Windows)	7085

“Seasonality” of MS Office XML Files

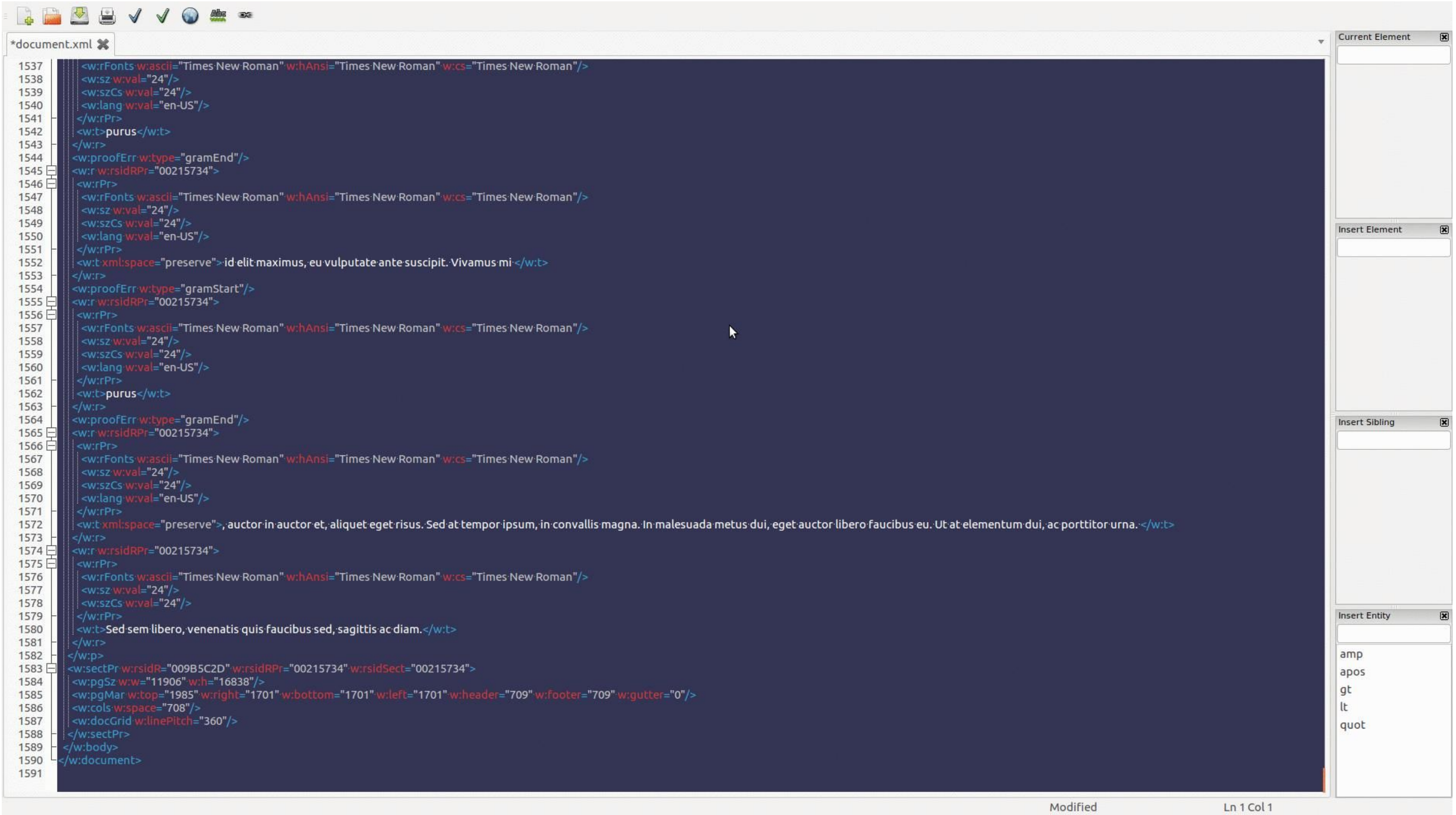
Versione	Lines of XML
Windows OOXML 2013 Transitional Summer 2017	1590
Windows OOXML 2013 Transitional Winter 2018	13515
Windows OOXML 2016 Transitional Summer 2017	11667
Windows OOXML 2016 Transitional Winter 2018	969
Windows OOXML 2016 Transitional Fall 2018	11288
Windows OOXML 2016 Transitional Spring 2019	7085
MacOS OOXML 2016 Transitional Summer 2017	11646
MacOS OOXML 2016 Transitional Fall 2018	854
MacOS OOXML 2016 Transitional Spring 2019	7731

XML of ODT (LibreOffice)



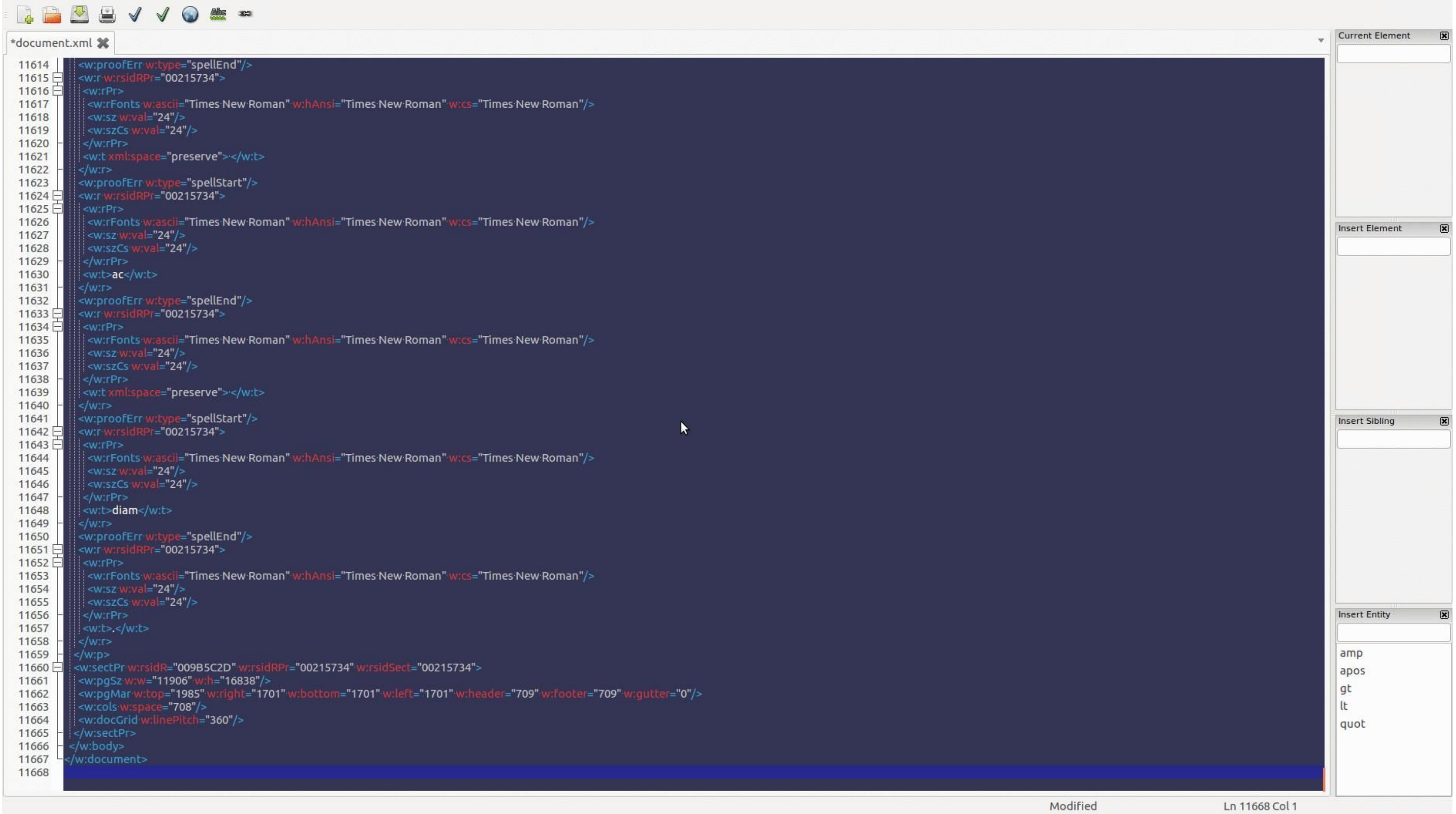
```xml
   <table:table table:name="Table1" table:style-name="Table1">
    <table:table-column table:style-name="Table1.A" table:number-columns-repeated="2"/>
    <table:table-row>
     <table:table-cell table:style-name="Table1.A1" office:value-type="string">
      <text:p text:style-name="P8">Aenean ultricies turpis eu erat euismod porta.</text:p>
     </table:table-cell>
     <table:table-cell table:style-name="Table1.B1" office:value-type="string">
      <text:p text:style-name="P8">Maecenas nec ullamcorper odio, ac tincidunt metus.</text:p>
     </table:table-cell>
    </table:table-row>
    <table:table-row>
     <table:table-cell table:style-name="Table1.A2" office:value-type="string">
      <text:p text:style-name="P8">Duis volutpat sapien at interdum consequat.</text:p>
     </table:table-cell>
     <table:table-cell table:style-name="Table1.B2" office:value-type="string">
      <text:p text:style-name="P8">Donec vel semper nisl.</text:p>
     </table:table-cell>
    </table:table-row>
    <table:table-row>
     <table:table-cell table:style-name="Table1.A2" office:value-type="string">
      <text:p text:style-name="P8">Sed scelerisque odio non justo suscipit congue.</text:p>
     </table:table-cell>
     <table:table-cell table:style-name="Table1.B2" office:value-type="string">
      <text:p text:style-name="P8">Maecenas convallis quis neque at elementum.</text:p>
     </table:table-cell>
    </table:table-row>
    <table:table-row>
     <table:table-cell table:style-name="Table1.B2" table:number-columns-spanned="2" office:value-type="string">
      <text:p text:style-name="P8">In fringilla tortor id orci porttitor, a euismod velit molestie.</text:p>
     </table:table-cell>
     <table:covered-table-cell/>
    </table:table-row>
   </table:table>
   <text:p text:style-name="P5"/>
   <text:p text:style-name="P3">Sed vel enim molestie, viverra sapien in, lobortis erat.</text:p>
   <text:p text:style-name="P7"/>
   <text:p text:style-name="P7">Suspendisse at mauris eget mi tempus dapibus vel in ante. Vivamus euismod, nisl commodo pharetra molestie, erat est elementum orci, ullamcorper bibendum lorem dui porttitor lorem. Phasellus tristique justo ut arcu tempor dapibus. Cras massa leo, dignissim sit amet mattis in, eleifend quis quam. Duis in massa ipsum. Quisque sed nibh ac sapien ornare hendrerit a non turpis. Aenean auctor, massa eget elementum auctor, odio dolor mattis felis, ut cursus augue magna vel elit. Quisque sit amet felis felis. Sed condimentum nulla ac felis viverra suscipit. Nulla lorem justo, dictum at facilisis nec, commodo non eros. Aliquam sapien sem, maximus ut lacus sed, lobortis convallis turpis. Donec tincidunt elit lacus, ut bibendum odio feugiat et.</text:p>
   <text:p text:style-name="P5"/>
   <text:p text:style-name="P5">Maecenas eleifend accumsan lacus, in elementum sapien sollicitudin non. Etiam a augue sed metus viverra gravida. Nullam interdum felis nec lectus congue ultrices. Nunc lobortis ex eu lobortis vehicula. Orci varius natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aenean felis nisl, auctor ac mauris sit amet, sodales tristique sapien. Curabitur auctor mattis orci quis gravida. Morbi quis orci enim. Nulla porta tortor in interdum efficitur. Nullam ac laoreet orci.</text:p>
   <text:p text:style-name="P5"/>
   <text:p text:style-name="P4">Aliquam lobortis rutrum massa.</text:p>
   <text:p text:style-name="P9"/>
   <text:p text:style-name="P9">Aenean iaculis tortor libero, at scelerisque ante porttitor vitae. Nulla pulvinar eget ipsum nec cursus. Proin quis mauris eu libero dictum tincidunt. Aliquam tristique dolor vel mi suscipit rutrum. Donec lorem tellus, aliquam ac risus vitae, congue ultricies tortor. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam ut blandit ante. Pellentesque consequat risus feugiat tortor tempus, cursus porttitor ipsum consequat. Pellentesque blandit sit amet erat a dictum. Donec scelerisque purus id elit maximus, eu vulputate ante suscipit. Vivamus mi purus, auctor in auctor et, aliquet eget risus. Sed at tempor ipsum, in convallis magna. In malesuada metus dui, eget auctor libero faucibus eu. Ut at elementum dui, ac porttitor urna. Sed sem libero, venenatis quis faucibus sed, sagittis ac diam.</text:p>
  </office:text>
 </office:body>
</office:document-content>
```

XML of DOCX (MS Office 2013)



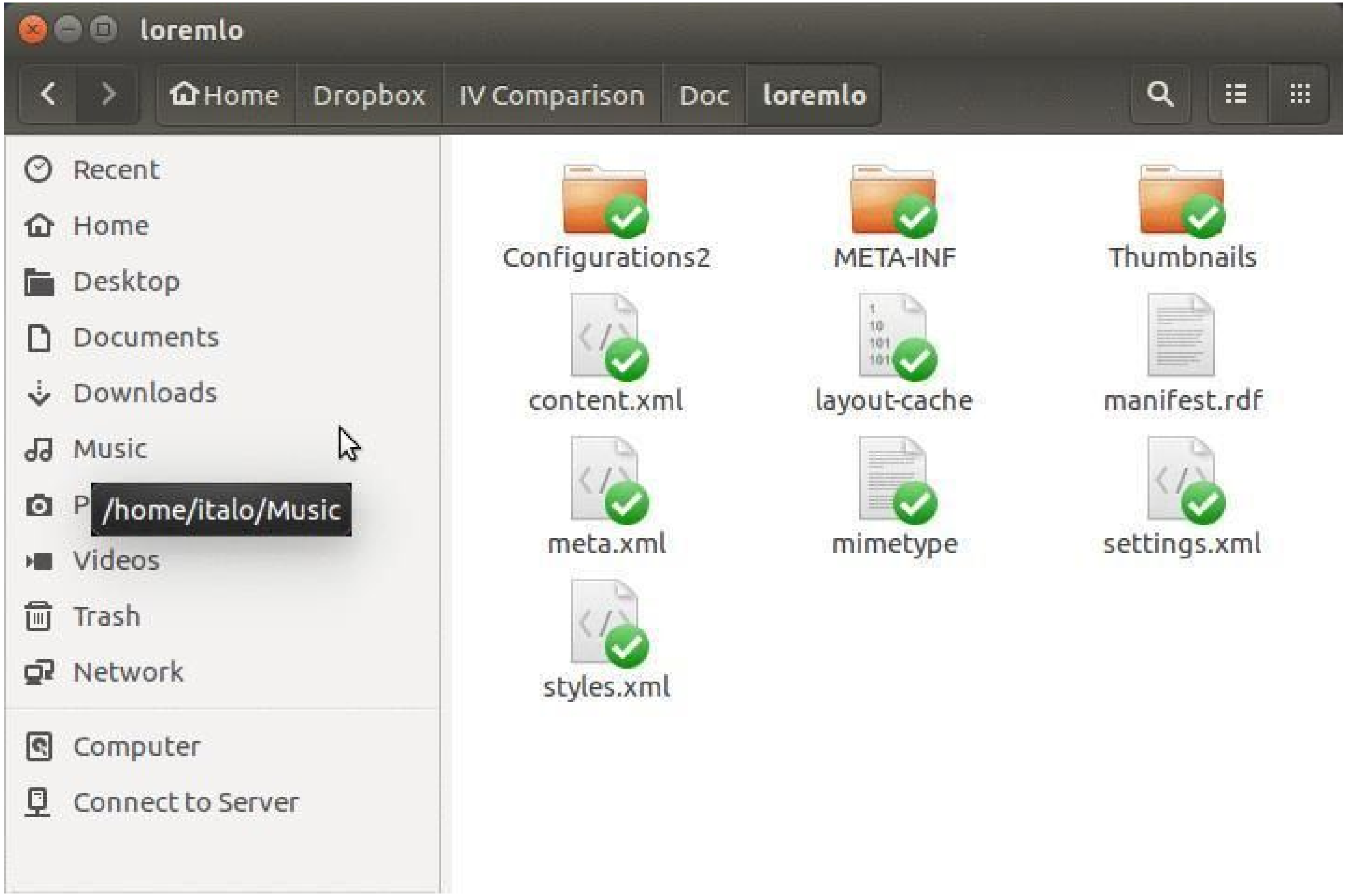
```
          <w:rFonts w:ascii="Times New Roman" w:hAnsi="Times New Roman" w:cs="Times New Roman"/>
          <w:sz w:val="24"/>
          <w:szCs w:val="24"/>
          <w:lang w:val="en-US"/>
        </w:rPr>
        <w:t>purus</w:t>
      </w:r>
      <w:proofErr w:type="gramEnd"/>
      <w:r w:rsidRPr="00215734">
        <w:rPr>
          <w:rFonts w:ascii="Times New Roman" w:hAnsi="Times New Roman" w:cs="Times New Roman"/>
          <w:sz w:val="24"/>
          <w:szCs w:val="24"/>
          <w:lang w:val="en-US"/>
        </w:rPr>
        <w:t xml:space="preserve"> id elit maximus, eu vulputate ante suscipit. Vivamus mi </w:t>
      </w:r>
      <w:proofErr w:type="gramStart"/>
      <w:r w:rsidRPr="00215734">
        <w:rPr>
          <w:rFonts w:ascii="Times New Roman" w:hAnsi="Times New Roman" w:cs="Times New Roman"/>
          <w:sz w:val="24"/>
          <w:szCs w:val="24"/>
          <w:lang w:val="en-US"/>
        </w:rPr>
        <w:t>purus</w:t>
      </w:r>
      <w:proofErr w:type="gramEnd"/>
      <w:r w:rsidRPr="00215734">
        <w:rPr>
          <w:rFonts w:ascii="Times New Roman" w:hAnsi="Times New Roman" w:cs="Times New Roman"/>
          <w:sz w:val="24"/>
          <w:szCs w:val="24"/>
          <w:lang w:val="en-US"/>
        </w:rPr>
        <w:t xml:space="preserve">, auctor in auctor et, aliquet eget risus. Sed at tempor ipsum, in convallis magna. In malesuada metus dui, eget auctor libero faucibus eu. Ut at elementum dui, ac porttitor urna. </w:t>
      </w:r>
      <w:r w:rsidRPr="00215734">
        <w:rPr>
          <w:rFonts w:ascii="Times New Roman" w:hAnsi="Times New Roman" w:cs="Times New Roman"/>
          <w:sz w:val="24"/>
          <w:szCs w:val="24"/>
        </w:rPr>
        <w:t>Sed sem libero, venenatis quis faucibus sed, sagittis ac diam.</w:t>
      </w:r>
    </w:p>
    <w:sectPr w:rsidR="009B5C2D" w:rsidRPr="00215734" w:rsidSect="00215734">
      <w:pgSz w:w="11906" w:h="16838"/>
      <w:pgMar w:top="1985" w:right="1701" w:bottom="1701" w:left="1701" w:header="709" w:footer="709" w:gutter="0"/>
      <w:cols w:space="708"/>
      <w:docGrid w:linePitch="360"/>
    </w:sectPr>
  </w:body>
</w:document>
```

XML of DOCX (MS Office 2016)



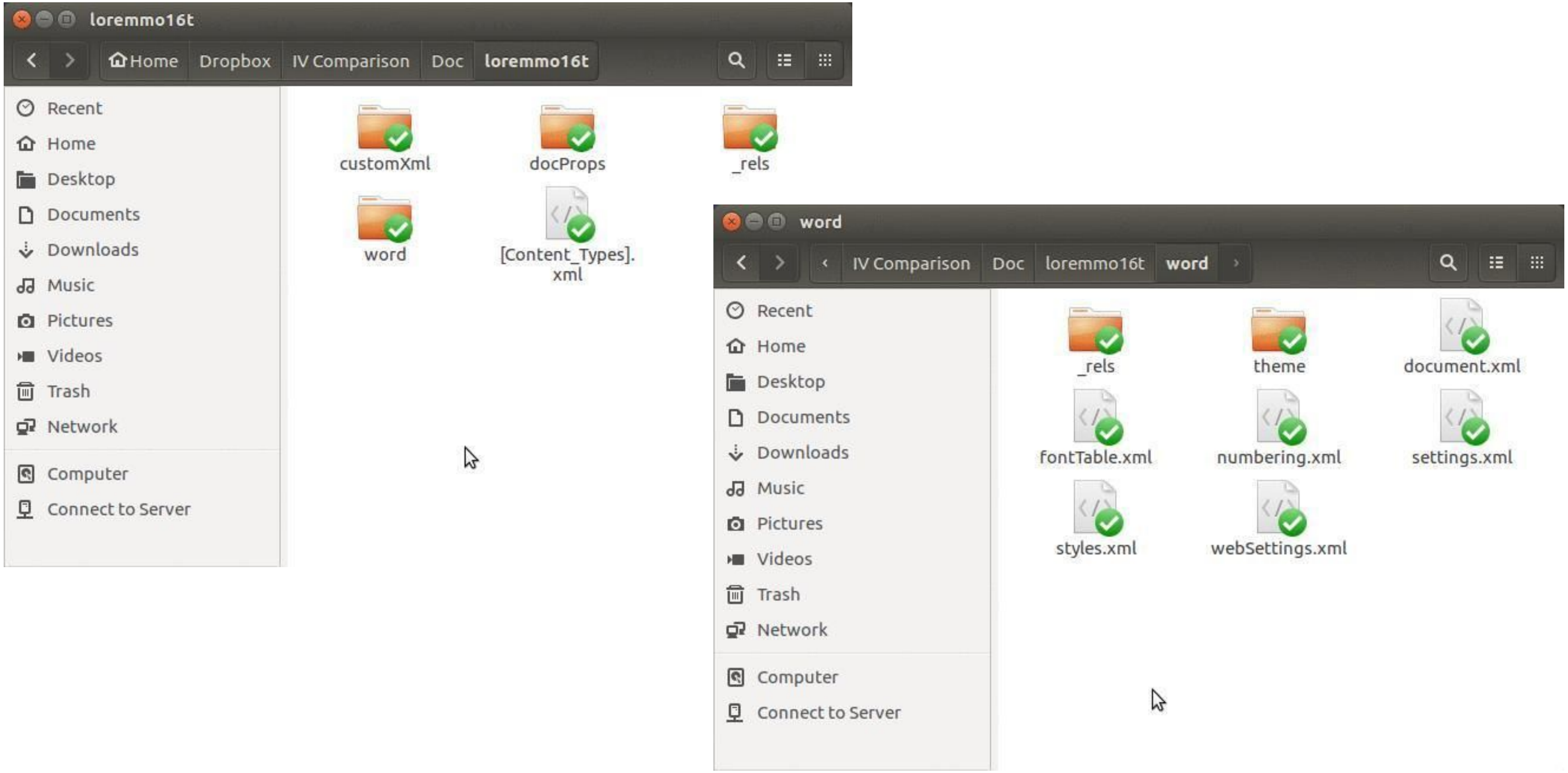
```
    <w:proofErr w:type="spellEnd"/>
    <w:r w:rsidRPr="00215734">
     <w:rPr>
      <w:rFonts w:ascii="Times New Roman" w:hAnsi="Times New Roman" w:cs="Times New Roman"/>
      <w:sz w:val="24"/>
      <w:szCs w:val="24"/>
     </w:rPr>
     <w:t xml:space="preserve"> </w:t>
    </w:r>
    <w:proofErr w:type="spellStart"/>
    <w:r w:rsidRPr="00215734">
     <w:rPr>
      <w:rFonts w:ascii="Times New Roman" w:hAnsi="Times New Roman" w:cs="Times New Roman"/>
      <w:sz w:val="24"/>
      <w:szCs w:val="24"/>
     </w:rPr>
     <w:t>ac</w:t>
    </w:r>
    <w:proofErr w:type="spellEnd"/>
    <w:r w:rsidRPr="00215734">
     <w:rPr>
      <w:rFonts w:ascii="Times New Roman" w:hAnsi="Times New Roman" w:cs="Times New Roman"/>
      <w:sz w:val="24"/>
      <w:szCs w:val="24"/>
     </w:rPr>
     <w:t xml:space="preserve"> </w:t>
    </w:r>
    <w:proofErr w:type="spellStart"/>
    <w:r w:rsidRPr="00215734">
     <w:rPr>
      <w:rFonts w:ascii="Times New Roman" w:hAnsi="Times New Roman" w:cs="Times New Roman"/>
      <w:sz w:val="24"/>
      <w:szCs w:val="24"/>
     </w:rPr>
     <w:t>diam</w:t>
    </w:r>
    <w:proofErr w:type="spellEnd"/>
    <w:r w:rsidRPr="00215734">
     <w:rPr>
      <w:rFonts w:ascii="Times New Roman" w:hAnsi="Times New Roman" w:cs="Times New Roman"/>
      <w:sz w:val="24"/>
      <w:szCs w:val="24"/>
     </w:rPr>
     <w:t>.</w:t>
    </w:r>
   </w:p>
   <w:sectPr w:rsidR="009B5C2D" w:rsidRPr="00215734" w:rsidSect="00215734">
    <w:pgSz w:w="11906" w:h="16838"/>
    <w:pgMar w:top="1985" w:right="1701" w:bottom="1701" w:left="1701" w:header="709" w:footer="709" w:gutter="0"/>
    <w:cols w:space="708"/>
    <w:docGrid w:linePitch="360"/>
   </w:sectPr>
  </w:body>
 </w:document>
```

Inside ODT

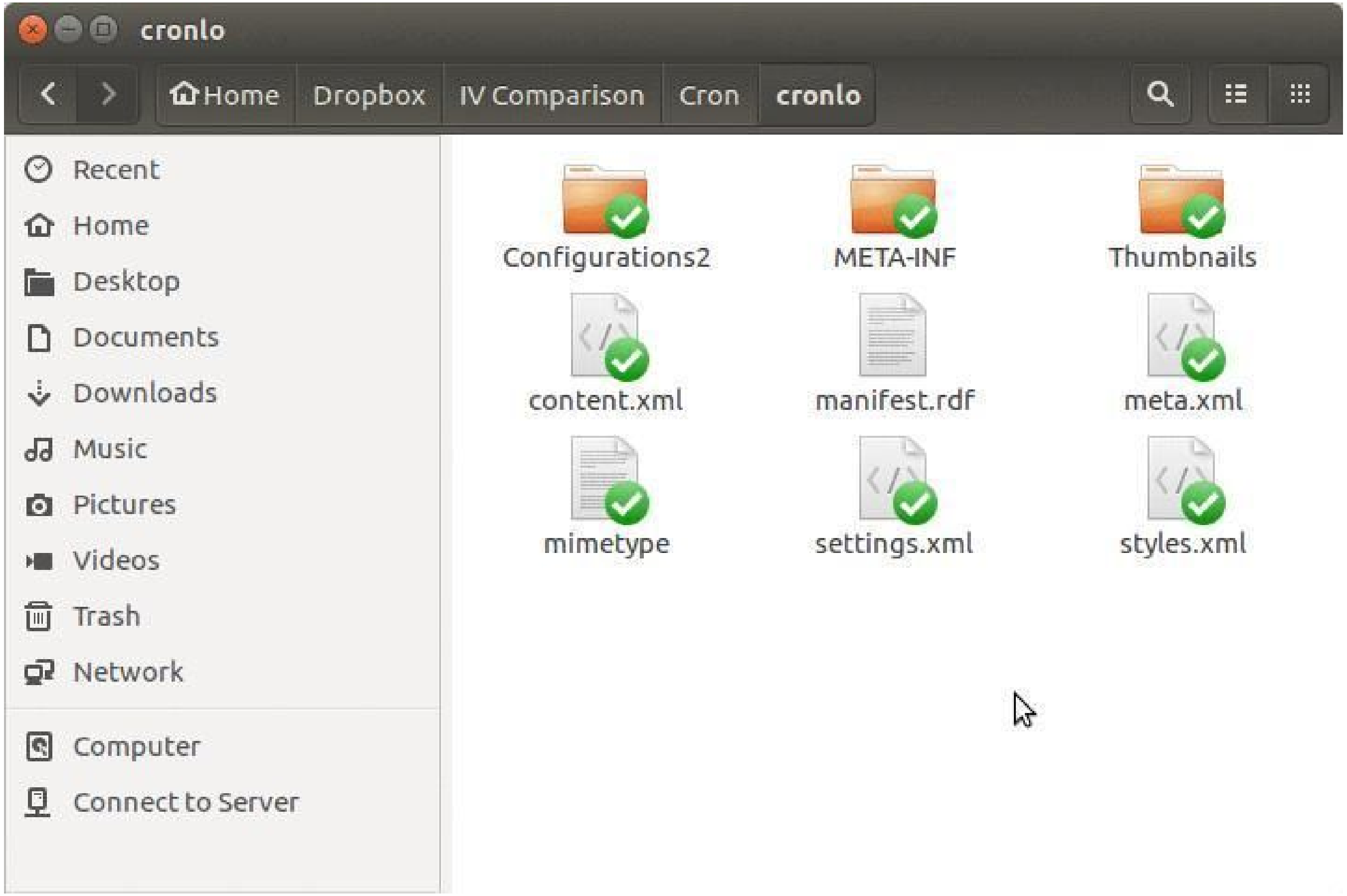


loremlo
Home
Dropbox
IV Comparison
Doc
loremlo
Recent
Home
Desktop
Documents
Downloads
Music
P
/home/italo/Music
Videos
Trash
Network
Computer
Connect to Server
Configurations2
META-INF
Thumbnails
content.xml
layout-cache
manifest.rdf
meta.xml
mimetype
settings.xml
styles.xml

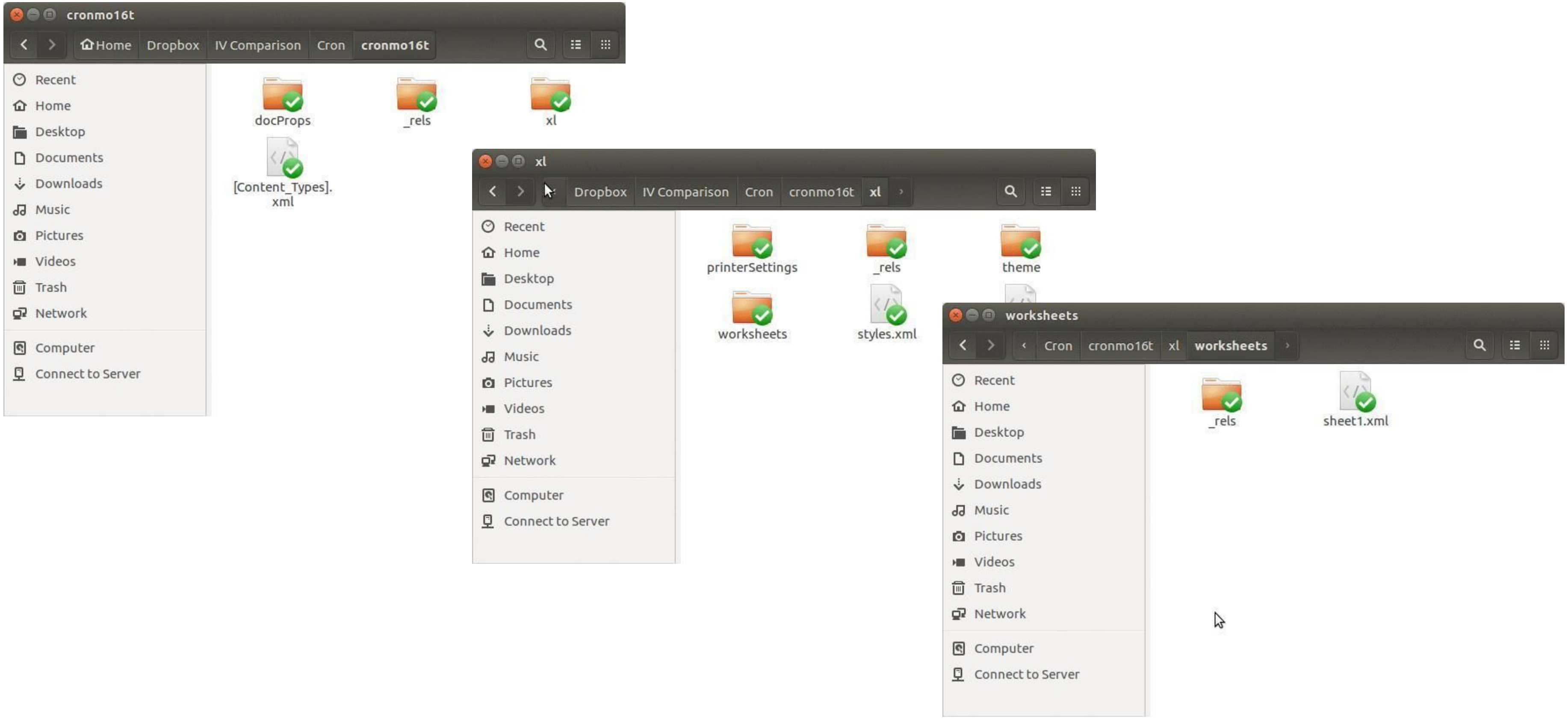
Inside DOCX


loremmo16t
Home
Dropbox
IV Comparison
Doc
loremmo16t
Recent
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Network
Computer
Connect to Server
customXml
docProps
_rels
word
[Content_Types].xml
word
IV Comparison
Doc
loremmo16t
word
Recent
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Network
Computer
Connect to Server
_rels
theme
document.xml
fontTable.xml
numbering.xml
settings.xml
styles.xml
webSettings.xml

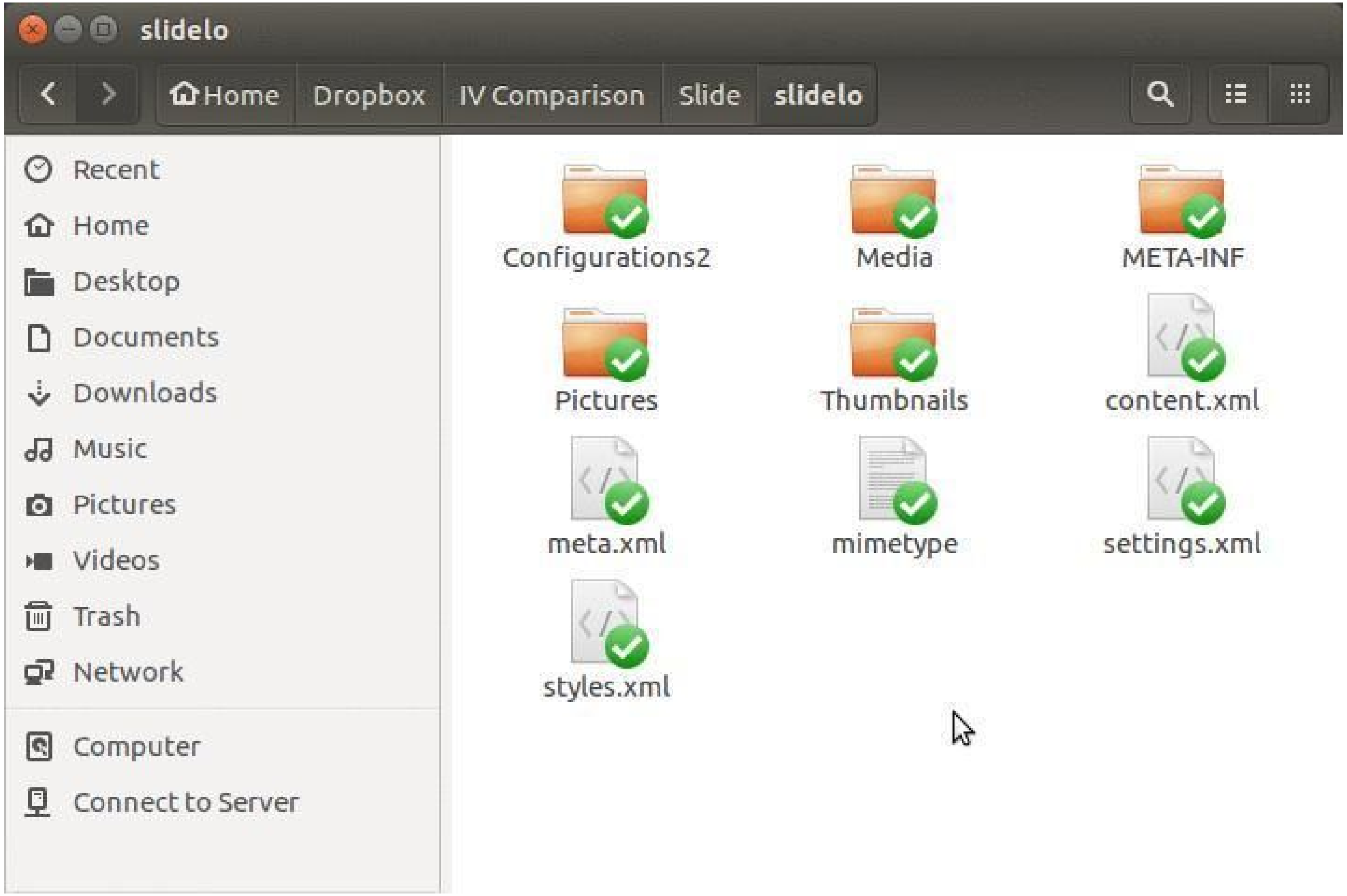
Inside ODS


cronlo
Home
Dropbox
IV Comparison
Cron
cronlo
Recent
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Network
Computer
Connect to Server
Configurations2
META-INF
Thumbnails
content.xml
manifest.rdf
meta.xml
mimetype
settings.xml
styles.xml

Inside XLSX

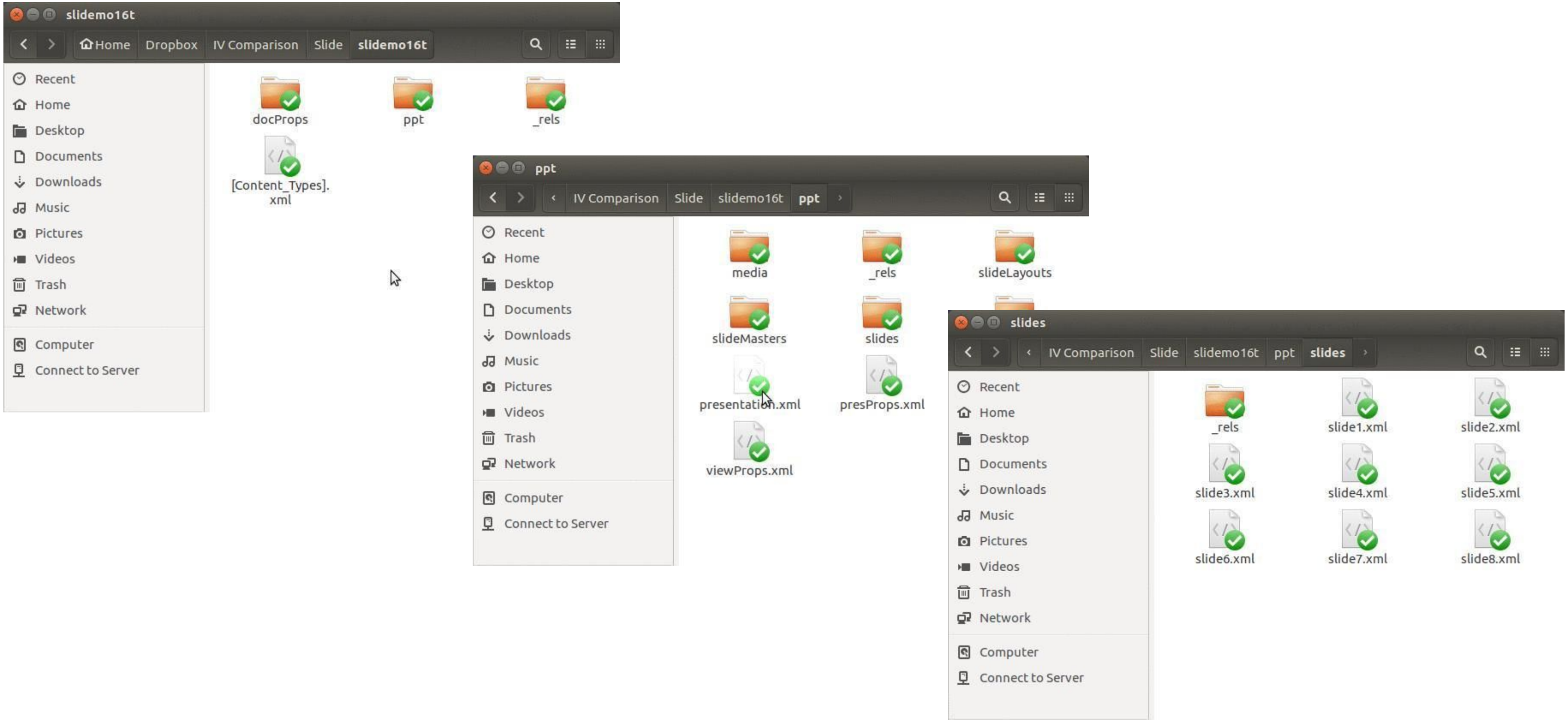

cronmo16t
Home
Dropbox
IV Comparison
Cron
cronmo16t
Recent
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Network
Computer
Connect to Server
docProps
_rels
xl
[Content_Types].xml
xl
Dropbox
IV Comparison
Cron
cronmo16t
xl
printerSettings
_rels
theme
worksheets
styles.xml
worksheets
Cron
cronmo16t
xl
worksheets
_rels
sheet1.xml

Inside ODP



Configurations2
Media
META-INF
Pictures
Thumbnails
content.xml
meta.xml
mimetype
settings.xml
styles.xml

Inside PPTX


slidemo16t
Home
Dropbox
IV Comparison
Slide
slidemo16t
Recent
Home
Desktop
Documents
Downloads
Music
Pictures
Videos
Trash
Network
Computer
Connect to Server
docProps
ppt
_rels
[Content_Types].xml
ppt
IV Comparison
Slide
slidemo16t
ppt
media
_rels
slideLayouts
slideMasters
slides
presentation.xml
presProps.xml
viewProps.xml
slides
IV Comparison
Slide
slidemo16t
ppt
slides
_rels
slide1.xml
slide2.xml
slide3.xml
slide4.xml
slide5.xml
slide6.xml
slide7.xml
slide8.xml

Simplicity vs Hidden Complexity

- **ODT / LibreOffice**
  - Reduced, very low or non existing complexity
  - XML files are human readable (as they should be)
- **OOXML / Microsoft Office**
  - Highest possible complexity vs technology
  - XML files are not human readable (contrary to what the XML standard language mandates)

Myth: OOXML is Better Documented

- No, more than 80% of OOXML huge documentation (over 7K pages) is used to "reinvent the wheel":
  - Describe proprietary Microsoft formats adopted to replace available open standards (i.e, VML over SVG)
  - Describe OOXML incredibly complex XML Schema, which is not following any XML convention (i.e., text for "text", strong for "bold", etc.)
  - Describe proprietary elements of legacy MS Office formats, which are not part of the ISO standard

Myth: OOXML is Backward Compatible

- No, the truth is exactly the opposite:
  - OOXML is backward compatible with legacy proprietary MS Office formats, but not with OOXML standard documents (which do not exist as OOXML Strict has never been adopted by users)
  - OOXML files are not versioned, are not consistent over time (probably because of different versions, if they were documented), and 100% of them is still proprietary OOXML Transitional

Thanks

Italo Vignoli
Chair, ODF Advocacy
italo@libreoffice.org
+39 348 5653829