Open Formats
ODF vs OOXML

Italo Vignoli
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
Definition of Open Standard

- Promotes a healthy competitive market (the existence of Open Standards reduces the risk and cost of market entry, and so encourages multiple suppliers)
- Reduces the risk to an organisation of being technologically locked-in
- Is a basis for interoperability, which supports systems heterogeneity, thereby increasing options for organisations
- Offers a basis for long-term access and reuse of digital assets, and in particular when supported by Open Source Reference Implementations
FOSS and Open Standards

- Support open standards wherever possible
- When given an alternative, prefer the most open standard that solves the problem
- Use open standards in every project activity
- Get involved in standards committees
- Help to develop and promote new standards
Open Format

• **Independent** from a single product: anyone can write a software that handles an open format
• **Interoperable**: allows the transparent sharing of data between heterogeneous systems
• **Neutral**: it does not force the user to adopt – and often buy – a specific product, but leaves a wide choice based on quality/price ratio
• **Perennial**: protects user developed contents from the “evolution” based obsolescence of technology
Proprietary Format

- Designed to be manipulated by a single software
- Evolves over the years based on commercial strategies and not on user needs
- Often, a direct serialization of data structures in memory
- The software is the format!
- Users borrow content from vendors through the format End User License Agreement (EULA)
Lock In

WE CANNOT READ YOUR DOCUMENTS

DOCUMENTFREEDOM.ORG
How to Lock-in Your Clients
How Professional Services Firms Can Create Compelling Value for Clients Using Collaborative Technologies

Ross Dawson
CEO, Advanced Human Technologies
Author, Living Networks and Developing Knowledge-Based Client Relationships

January 2004
Interoperability
Interoperability means the ability of information and communication technology (ICT) systems, as well as, of the business processes they support in order to exchange data and enable the sharing of information and knowledge.

*European Interoperability Framework, IDABC*
Legacy Interoperability
Benefits of Interoperability

- **FINANCIAL**
  - Cost savings for users
  - Reduced operation costs for owners
  - Vendor lock-in avoidance for owners
  - Facilitate reuse, sharing & adoption

- **TIME**
  - Owner time savings
  - User time savings

- **QUALITY**
  - High service satisfaction for users
  - Improved compliance for owners
  - Better data quality for owners
  - Better data availability for users
  - Improved security for owners

- **OTHER**
  - Foster innovation
  - Increase transparency
  - Protection of user’s rights
  - Furthering public policy goals
  - Reduced CO2 emissions
Interoperability Costs (1)

Cost Analysis of Inadequate Interoperability in the U.S. Capital Facilities Industry

Michael P. Gallaher, Alan C. O’Connor, John L. Dettbarn, Jr., and Linda T. Gilday
## Interoperability Costs (2)

### Table 6-5. Costs of Inadequate Interoperability for Architects and Engineers

<table>
<thead>
<tr>
<th>Life-Cycle Phase</th>
<th>Cost Category</th>
<th>Cost Component</th>
<th>Average Cost per Square Foot</th>
<th>Average Cost per Square Meter</th>
<th>Inadequate Interoperability Cost Estimate ($Thousands)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning, Engineering, and Design</td>
<td></td>
<td>Inefficient business process management costs</td>
<td>0.31</td>
<td>3.37</td>
<td>356,126</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Redundant CAx systems costs</td>
<td>0.0001</td>
<td>0.001</td>
<td>158</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Productivity losses and training costs for redundant CAx systems</td>
<td>0.04</td>
<td>0.45</td>
<td>47,947</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Redundant IT support staffing for CAx systems</td>
<td>0.0004</td>
<td>0.005</td>
<td>501</td>
</tr>
<tr>
<td>Avoidance Costs</td>
<td></td>
<td>Data translation costs</td>
<td>0.002</td>
<td>0.02</td>
<td>2,139</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Interoperability research and development expenditures</td>
<td>0.02</td>
<td>0.21</td>
<td>22,234</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Manual reentry costs</td>
<td>0.41</td>
<td>4.38</td>
<td>462,734</td>
</tr>
<tr>
<td>Mitigation Costs</td>
<td></td>
<td>Design and construction information verification costs</td>
<td>0.10</td>
<td>1.08</td>
<td>114,342</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Reworking design files costs</td>
<td>0.0009</td>
<td>0.009</td>
<td>968</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Avoidance costs</td>
<td>0.38</td>
<td>3.85</td>
<td>429,106</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mitigation costs</td>
<td>0.51</td>
<td>5.47</td>
<td>578,044</td>
</tr>
<tr>
<td>Subtotal</td>
<td></td>
<td>Subtotal</td>
<td>0.89</td>
<td>9.32</td>
<td>1,007,150</td>
</tr>
</tbody>
</table>
# Interoperability Costs (3)

<table>
<thead>
<tr>
<th>Construction Costs</th>
<th>Avoidance Costs</th>
<th>Inefficient business process management costs</th>
<th>0.04</th>
<th>0.41</th>
<th>43,290</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Redundant CAx systems costs</td>
<td>0.00003</td>
<td>0.0003</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Productivity losses and training costs</td>
<td>0.007</td>
<td>0.08</td>
<td>8,461</td>
</tr>
<tr>
<td></td>
<td></td>
<td>for redundant CAx systems</td>
<td>0.00008</td>
<td>0.0008</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Data translation costs</td>
<td>0.0003</td>
<td>0.004</td>
<td>378</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Interoperability research and development</td>
<td>0.003</td>
<td>0.04</td>
<td>3,924</td>
</tr>
<tr>
<td></td>
<td></td>
<td>expenditures</td>
<td>0.024</td>
<td>0.26</td>
<td>27,750</td>
</tr>
<tr>
<td></td>
<td>Mitigation Costs</td>
<td>Design and construction information</td>
<td>0.006</td>
<td>0.07</td>
<td>7,377</td>
</tr>
<tr>
<td></td>
<td></td>
<td>verification costs</td>
<td>0.05</td>
<td>0.53</td>
<td>55,656</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RFI management costs</td>
<td>0.05</td>
<td>0.53</td>
<td>55,656</td>
</tr>
<tr>
<td></td>
<td><strong>Subtotal</strong></td>
<td>Avoidance costs</td>
<td>0.05</td>
<td>0.49</td>
<td>56,169</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mitigation costs</td>
<td>0.08</td>
<td>0.86</td>
<td>90,783</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Subtotal</strong></td>
<td><strong>0.13</strong></td>
<td><strong>1.35</strong></td>
<td><strong>146,952</strong></td>
</tr>
<tr>
<td>Operations and</td>
<td>Mitigation</td>
<td>Post-construction redundant information</td>
<td>0.01</td>
<td>0.15</td>
<td>15,660</td>
</tr>
<tr>
<td>Maintenance</td>
<td>Costs</td>
<td>transfer costs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Subtotal</strong></td>
<td>Total Cost</td>
<td><strong>1.169,762</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: RTI estimates; totals may not sum correctly due to rounding.
Summary of Interoperability

TECHNOLOGICAL
hardware and code to allow connections

DATA
ability of interconnected systems to understand each other

INSTITUTIONAL
effective engagement of societal systems

HUMAN
ability to understand and act on data exchanged
Openness & Interoperability
Open Document Format
the true document standard
which offers freedom of choice
ODF is an Open Standard

- Open and collaborative development
- Transparent access to the minutes of all meetings
- Ease of implementation in any software
- Freedom from patents and licensing restrictions
- No reliance on proprietary features or single vendor owned technologies
- Interoperability with any system designed to be ODF compatible
What is ODF? (1)

- OpenDocument Format (ISO / IEC 26300)
  - ISO: International Organization for Standardization
  - IEC: International Electrotechnical Commission
- Open file format based on XML, to create, view, edit and store Office documents
- Text documents, spreadsheets, and presentations
What is ODF? (2)

- Defined with an open and transparent process by OASIS
- Approved by the Joint Technical Committee 1 (JTC 1) of the IEC as International Standard (IS) in May 2006
- Available for deployment and use with no license, royalty payments or other restrictions
ODF Based Interoperability

OLD STYLE
Content closely related to a specific application
Developers and not users control the application

NEW STYLE
Content represented through an open standard which is not controlled by a single vendor, and supported by many applications
Users are in full control of their contents
ODF is Standard

- France
- Portugal
- Sweden
- Taiwan
- UK
## ODF Document Extensions

<table>
<thead>
<tr>
<th>Document</th>
<th>Extension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>ODT</td>
</tr>
<tr>
<td>Spreadsheet</td>
<td>ODS</td>
</tr>
<tr>
<td>Presentation</td>
<td>ODP</td>
</tr>
<tr>
<td>Graphics</td>
<td>ODG</td>
</tr>
<tr>
<td>Business Chart</td>
<td>ODC</td>
</tr>
<tr>
<td>Image</td>
<td>ODI</td>
</tr>
<tr>
<td>Database</td>
<td>ODB</td>
</tr>
<tr>
<td>Math Formula</td>
<td>ODF</td>
</tr>
</tbody>
</table>
Inside ODT
Inside ODS
Inside ODP
Free Fonts
ODF vs OOXML
Battle of 2 Standards
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: a pseudo-standard

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

Office Open XML Overview
ECMA International, December 2006
Document Complexity

- **OBJECTS** - Typography, bitmap and outline images, colour, business rules, text, steganography, and much more are used to make up documents, however the management and approval process of these objects are also key to ensuring the correctness, and value of each individual document.

- **DATA** - You need to understand the structure of data, storage locations and techniques, extraction, backup, transmission methods, normalisation, consolidation, translation, manipulation, and sorting, plus security, privacy, and data governance.
Complexity of IT Systems

complexity

time
Hidden Complexity

complexity

hidden complexity

visible complexity

time
Reuse of Existing Standards

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

OOXML
- Dublin Core
Brain and Computer

- Brain
- Computer

Red

#FF0000
Description of Colours

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"
Creative Handling of Calendar

OOXML
Microsoft Office

ODF
LibreOffice
# Real Dates vs Excel Dates

<table>
<thead>
<tr>
<th>Event</th>
<th>Calc</th>
<th>Excel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Italo Vignoli Birthday</td>
<td>12/08/1954</td>
<td>19948</td>
</tr>
<tr>
<td>Italo Vignoli Graduation</td>
<td>19/11/1978</td>
<td>28813</td>
</tr>
<tr>
<td>Italo Vignoli First Job</td>
<td>01/10/1981</td>
<td>29860</td>
</tr>
<tr>
<td>Italo Vignoli First PC</td>
<td>01/09/1983</td>
<td>30560</td>
</tr>
<tr>
<td>Italo Vignoli Wedding</td>
<td>08/09/1984</td>
<td>30933</td>
</tr>
<tr>
<td>Italo Vignoli Installs OOo</td>
<td>02/01/2003</td>
<td>37623</td>
</tr>
<tr>
<td>Italo Vignoli Launches LibreOffice</td>
<td>28/09/2010</td>
<td>40449</td>
</tr>
</tbody>
</table>
Other OOXML Issues

- OOXML does not consider Jewish and Muslim users, who use a different calendar.
- OOXML uses its own proprietary list of numbers instead of the ISO 639 standard for language names (English = EN).
- OOXML adopts the VML vector format, which conflicts with ISO/IEC 8632/W3C SVG standard (VML was deprecated by Microsoft before the development of OOXML).
- OOXML adopts Math for mathematical notation instead of W3C MathML (Math is intentionally incompatible with the standard MathML).
File Comparison
ODF vs OOXML
Comparison ODF/OOXML
# Length of Content XML

<table>
<thead>
<tr>
<th>Version</th>
<th>XML Lines</th>
</tr>
</thead>
<tbody>
<tr>
<td>ODF 1.2 (any version of) LibreOffice</td>
<td>222</td>
</tr>
<tr>
<td>OOXML 2010 Transitional (MS Office Windows)</td>
<td>1040</td>
</tr>
<tr>
<td>OOXML 2011 Transitional (MS Office MacOS)</td>
<td>12854</td>
</tr>
<tr>
<td>OOXML 2013 Transitional (MS Office Windows)</td>
<td>1590</td>
</tr>
<tr>
<td>OOXML 2016 Transitional (MS Office Windows)</td>
<td>11667</td>
</tr>
<tr>
<td>OOXML 2016 Transitional (MS Office MacOS)</td>
<td>11646</td>
</tr>
<tr>
<td>OOXML 2019 Transitional (MS Office Windows)</td>
<td>7085</td>
</tr>
</tbody>
</table>
## Seasonality of Content XML

<table>
<thead>
<tr>
<th>Version</th>
<th>XML Lines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows OOXML 2013 Transitional Summer 2017</td>
<td>1590</td>
</tr>
<tr>
<td>Windows OOXML 2013 Transitional Winter 2018</td>
<td>13515</td>
</tr>
<tr>
<td>Windows OOXML 2016 Transitional Summer 2017</td>
<td>11667</td>
</tr>
<tr>
<td>Windows OOXML 2016 Transitional Winter 2018</td>
<td>969</td>
</tr>
<tr>
<td>Windows OOXML 2016 Transitional Fall 2018</td>
<td>11288</td>
</tr>
<tr>
<td>Windows OOXML 2016 Transitional Spring 2019</td>
<td>7085</td>
</tr>
<tr>
<td>MacOS OOXML 2016 Transitional Summer 2017</td>
<td>11646</td>
</tr>
<tr>
<td>MacOS OOXML 2016 Transitional Fall 2018</td>
<td>854</td>
</tr>
<tr>
<td>MacOS OOXML 2016 Transitional Spring 2019</td>
<td>7731</td>
</tr>
</tbody>
</table>
To be, or not to be, this is the question.
To be, or not to be, this is the problem.
Simplicity vs Hidden Complexity

- ODT by LibreOffice
  - Low or no hidden complexity
  - Same approach when writing OOXML
  - Files are human readable (security)
- OOXML by Microsoft Office
  - Highest option of hidden complexity
  - Same approach when writing ODT
  - Files are not human readable
Inside ODT
Inside DOCX
Inside ODS
Inside XLSX
Comparison Presentation
Inside ODP

Inside the ODP folder, there are several subfolders:
- Configurations2
- Media
- META-INF
  - Thumbnails
  - content.xml
- Pictures
  - meta.xml
  - mimetype
- Videos
  - styles.xml
- Documents
- Downloads
- Desktop
- Music
- Home
- Trash
- Network
- Computer
  - Connect to Server
Inside PPTX
Deductions of a Stupid Me

- LibreOffice developers are a bunch of geniuses
- Microsoft Office developers are a bunch of *****
  unless
- Microsoft Office XML files are artificially stuffed with useless contents to reduce the chances that software other than Microsoft Office can open them properly
- Microsoft has a vested interest in killing standard based interoperability to protect a captive market still valued at over 25 billion dollars
### Complexity of Document Formats

<table>
<thead>
<tr>
<th>Document Format</th>
<th># Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Office Open XML</td>
<td>1792</td>
</tr>
<tr>
<td>WordprocessingML</td>
<td>780</td>
</tr>
<tr>
<td>OASIS Open Document</td>
<td>530</td>
</tr>
</tbody>
</table>

*AS OF 2006*
File Types Used in Attacks

Source: Symantec MessageLabs Intelligence, February 2011 Intelligence Report
File Types Used in Attacks

Source: Kaspersky Labs, Spring 2019 Worldwide Meeting
LibreOffice vs MS Office

Comparison of Vulnerabilities 2015/2017

Software

Documenti

MS Office
LibreOffice
Non Standard: Circa 2000 AC
No Interoperability
Non Standard: Circa 2000 DC
False Interoperability
Standard: Circa 2000 DC
True Interoperability
Proprietary Fonts

Calibri – regular, italic, bold, bold italic
Cambria – regular, italic, bold, bold italic
Candara – regular, italic, bold, bold italic
Consolas – reg, it, bd, bd it
Constantia – regular, italic, bold, bold italic
Corbel – regular, italic, bold, bold italic
Cariadings –
Except where otherwise noted, content on this presentation is licensed under a Creative Commons Attribution 4.0 International license.