OpenDocument format: from COSM to the ODF Advocacy Open Project

Marina Latini
marina.latini@libreoffice.org

Italo Vignoli
italo@libreoffice.org

ALMERÍA | 11 Sept. 2018
COSM (1)

- **Community of ODF Specification Maintainers**
- Project started by TDF in early 2018, based on the need of evolving the ODF standard beyond version ODF 1.2 to reflect the work of Technical Committee members
- Independent from LibreOffice
- Project housed at Public Software CIC (a UK Community Interest Company)
• First crowdfunded open standards project: an effective way to have essential standards maintained without needing a deep-pocketed international corporation behind them

• Seeded by TDF in early 2018, and then funded by Microsoft, Collabora, CIB, the UK Government & the European Commission

• Paying for an editor to draft the ODF 1.3 specifications using the work of the great Technical Committee contributors
• ODF 1.3 Committee Draft approved September 9, 2019, and submitted to OASIS as a standard

• We do not expect significant reactions at this point
  – We do expect some random typo catching

• ODF 1.3 Committee Draft should be ratified by OASIS by the end of 2019

• ODF 1.3 ISO Standard expected for the end of 2020
ODF Advocacy
Open Project @ OASIS
Welcome to the home of the ODF Advocacy Open Project.

The ODF Advocacy project works to create awareness and educate the world about the benefits of using the OpenDocument Format OASIS Standard (also published as ISO/IEC 26300).

ODF is an XML-based file format for personal productivity applications such as office suites, word processors, text/document editors, spreadsheets, and presentation software. Use of ODF guarantees access to your data forever, ensuring that data can be transferred between different computers and operating systems, without having to worry about vendor lock-in or license fees.

The ODF Advocacy project develops non-commercial, informational materials for a sustained communication campaign about the technical advantages and cost-savings of using standard-based document interoperability over proprietary formats.

ODF Advocacy is an OASIS Open Project.
• The objective of the OpenDocument Format (ODF) Advocacy Open Project at OASIS is to create awareness and educate the world about the benefits of using ODF.

• The group will develop advocacy materials for a communication campaign about the importance of standard-based document interoperability versus proprietary formats.
Purpose and Scope

- Educate users of personal productivity applications such as office suites and the likes, about the advantages of adopting the international, ISO-published, ODF standard
- Help users migrate from pseudo-standard formats to ODF to ensure interoperability and save costs

**Awareness**: Dedicated Website, Social Media Campaign, Media Outreach Campaign, Conference Program, Plugfests

**Education**: What is a Document Standard?, Importance of Document Standards, Advantages of Document Standards, De Jure vs De Facto Standards

**Marketing**: Impact of Standards on Productivity, Economic Value of Interoperability, Competitive Advantages of ODF, Standards & Innovation
Business Benefits

- Users of personal productivity software will benefit from the ODF awareness and education campaign, as they will learn about how true document interoperability can protect their data into the future and ensure freedom-of-choice in software.

- Large organizations deploying personal productivity software will discover the savings made possible by ODF and the associated advantage of increased security thanks to the reduced number of vulnerabilities affecting files based on standard file formats.
Basic Concept
• ODF is solid and robust
• ODF is consistent across OS
• ODF is truly interoperable
• ODF is predictable

• **ODF is a better standard file format for users**
Food for Thought

ODF Advocacy
Lock In

WE CANNOT READ YOUR DOCUMENTS

DOCUMENTFREEDOM.ORG
Open Document Format
the true document standard
which offers freedom of choice
ODF is Standard

- France
- Portugal
- Sweden
- Taiwan
- UK
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)
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
Interoperability
Elements 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
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
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
## 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>Inefficient business process management costs</td>
<td>0.31</td>
<td>3.37</td>
<td>356,126</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Redundant CAx systems costs</td>
<td>0.0001</td>
<td>0.001</td>
<td>158</td>
<td></td>
</tr>
<tr>
<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>
<td></td>
</tr>
<tr>
<td></td>
<td>Redundant IT support staffing for CAx systems</td>
<td>0.0004</td>
<td>0.005</td>
<td>501</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Data translation costs</td>
<td>0.002</td>
<td>0.02</td>
<td>2,139</td>
<td></td>
</tr>
<tr>
<td>Avoidance Costs</td>
<td>Interoperability research and development expenditures</td>
<td>0.02</td>
<td>0.21</td>
<td>22,234</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Manual reentry costs</td>
<td>0.41</td>
<td>4.38</td>
<td>462,734</td>
<td></td>
</tr>
<tr>
<td>Mitigation Costs</td>
<td>Design and construction information verification costs</td>
<td>0.10</td>
<td>1.08</td>
<td>114,342</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reworking design files costs</td>
<td>0.0009</td>
<td>0.009</td>
<td>968</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Avoidance costs</td>
<td>0.38</td>
<td>3.85</td>
<td>429,106</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mitigation costs</td>
<td>0.51</td>
<td>5.47</td>
<td>578,044</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Subtotal</td>
<td><strong>0.89</strong></td>
<td><strong>9.32</strong></td>
<td><strong>1,007,150</strong></td>
<td></td>
</tr>
</tbody>
</table>
### Interoperability Costs (3)

<table>
<thead>
<tr>
<th>Construction Costs</th>
<th>Avoidance Costs</th>
<th>Mitigation Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Inefficient business process management costs 0.04 0.41 43,290</td>
<td>Design and construction information verification costs 0.006 0.07 7,377</td>
</tr>
<tr>
<td></td>
<td>Redundant CAx systems costs 0.00003 0.0003 28</td>
<td>RFI management costs 0.05 0.53 55,656</td>
</tr>
<tr>
<td></td>
<td>Productivity losses and training costs for redundant CAx systems 0.007 0.08 8,461</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Redundant IT support staffing for CAx systems 0.00008 0.0008 88</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Data translation costs 0.0003 0.004 378</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Interoperability research and development expenditures 0.003 0.04 3,924</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Manual reentry costs 0.024 0.26 27,750</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Subtotal</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Avoidance costs 0.05 0.49 56,169</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mitigation costs 0.08 0.86 90,783</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Subtotal 0.13 1.35 146,952</td>
<td></td>
</tr>
<tr>
<td>Operations and Maintenance Costs</td>
<td>Mitigation Costs</td>
<td>Post-construction redundant information transfer costs 0.01 0.15 15,660</td>
</tr>
<tr>
<td>Total Cost</td>
<td></td>
<td>1,169,762</td>
</tr>
</tbody>
</table>

Source: RTI estimates; totals may not sum correctly due to rounding.
ODF vs OOOXML
Battle of 2 Standards
• **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.
Hidden Complexity

- hidden complexity
- visible complexity

complexity

time
Brain and Computer

Brain

Red

Computer

#FF0000
Description of Colours

**ODF** (LibreOffice)
- Writer
  
  \[\text{fo:color} = \#FF0000\]
- Calc
  
  \[\text{fo:color} = \#FF0000\]
- Impress
  
  \[\text{fo:color} = \#FF0000\]

**OOXML** (MS Office)
- Word
  
  \[\text{w:color w:val} = \text{FF0000}\]
- Excel
  
  \[\text{color rgb} = \text{FFFF0000}\]
- PowerPoint
  
  \[\text{a:srgbClr val} = \text{FF0000}\]
<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>
## 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>
<tr>
<td>Version</td>
<td>XML Lines</td>
</tr>
<tr>
<td>--------------------------------------------------------</td>
<td>-----------</td>
</tr>
<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
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

Targeted platforms by attacked users

2016 Q4
- Browsers: 45%
- PDF: 1%
- Java: 6%
- Adobe Flash: 13%
- Office: 16%
- Android: 19%

2018 Q4
- Browsers: 14%
- PDF: 0%
- Java: 3%
- Adobe Flash: 1%
- Android: 12%
- Office: 70%

Source: Kaspersky Labs, Spring 2019 Worldwide Meeting
Except where otherwise noted, content on this presentation is licensed under a Creative Commons Attribution 4.0 International license.