



Interoperability via OpenXML

Wolfgang Keber
DlaLOGIKa – Germany
wk@dialogika.de

A Few Words about DIaLOGIKa

- Based in Germany (Saarbrücken, near Luxembourg)
- Founded in 1982 (still owned by the same people)
- Currently ca. 60 employees
- Specialised in consulting and development projects for
 - Multinational institutions (e.g. EU institutions)
 - Telecom service providers
 - Banking sector (secure communication)
 - Industry

Overview of the Presentation

- OpenXML
 - Main features
 - Technical details
 - Advantages
- OpenXML and Interoperability
 - Binary Office formats
 - Open Document Format

OpenXML – The New MS Office Format

- Officially “Office Open XML” (OOXML)
- Comes with Office 2007, but not bound to it
- Available for
 - Word (docx)
 - PowerPoint (pptx)
 - Excel (xlsx)
- What’s new in OpenXML?
 - Based on XML
 - Specification freely available (but voluminous)
 - Can be used by everyone (“covenant not to sue”)

OpenXML – A Look under the Hood

- Zip package as the container
 - A document is one single file
 - Small file size due to compression
- Subfiles inside the container
 - Content
 - Layout information
 - Metadata
 - Embedded objects, e.g. images
 - Relationship and content type files



OpenXML – New Options

- OpenXML documents can be processed on any platform
 - Document creation
 - Information extraction
 - Metadata
 - Text
 - Layout
 - Document modification
 - Exchanging embedded pictures, e.g. logos



OpenXML – Some More Options

- Compatibility with binary Office formats
 - OpenXML has been designed to faithfully preserve the features of existing Office documents
 - Converters are available
- Extending OpenXML (Word)
 - Custom XML parts
 - Stored in separate subfiles in the package
 - Bound to content controls
 - Automatically populate these controls
 - Extract information from these controls

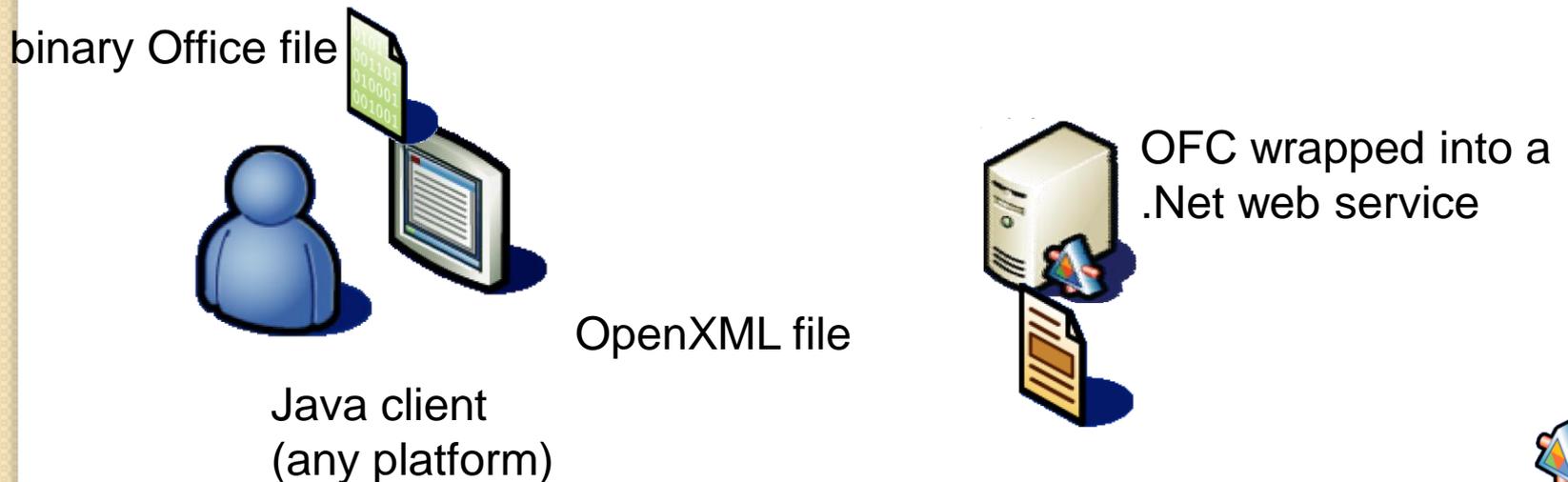


OpenXML and Binary Office Formats

- Office 2007
 - Opens/saves binary Office formats
 - Compatibility mode to preserve “old” functionality
- Office 2000-2003
 - File Format Compatibility Pack
 - Opens/saves OpenXML documents
- OMPM
 - Office Migration Planning Manager
 - Includes OFC (Office File Converter)
 - Converts binary files to OpenXML

OFC – Office File Converter

- Stand-alone conversion tool
 - Office applications are not required
- Can convert entire folder structures
- Could be wrapped into conversion services



OpenXML and the Open World

- OpenXML/ODF Translator Project
 - Open Source project hosted on SourceForge (<http://sourceforge.net/projects/odf-converter>)
 - Developed under a liberate BSD-like license
 - Several companies involved
 - Clever Age and Sonata (development)
 - Aztecsoft and DlaLOGIKa (testing)
 - Novell (Linux porting & OpenOffice.org integration)
 - Microsoft (funding and coordination)

ODF/OpenXML Translator (I)

- **Schedule**
 - Project start mid 2006
 - Version 1.0 for text documents (ODT/DOCX) released in January 2007
 - Translator for spreadsheets and presentations currently being developed
 - M2 released on 21 May
 - Version 1.1 for text documents in the pipeline
 - Version 1.0 user feedback taken into account
 - In particular feedback from EU (European Commission)
 - Final version planned for the end of August 2007

ODF/OpenXML Translator (2)

- Available in three variants
 - Add-in for Office 2002 and 2003 (+File Format Compatibility Pack)
 - Add-in for Office 2007
 - Command-line tool (Office apps not required)
- All use the same translation kernel
- Translation based on XSLT-technology
 - Pre- and postprocessing for special purposes
 - .Net Framework 2.0/C# for Office integration
 - Porting to other platforms feasible (see Novell/Linux)

It's a Translation (I)

- ODF and OpenXML implement partly different feature sets
 - One-to-one mapping
 - “bold” is mapped to “bold”
 - Workarounds or replacements
 - Intermediate font-weight values in ODF are translated to “normal” or “bold” in OpenXML
 - “SEQ” fields in OpenXML are translated to “Number range” fields in ODF
 - Section/page layout handling
 - Incompatible features
 - Picture rotation doesn't exist in ODF
 - Last line alignment in justified paragraph doesn't exist in OpenXML

It's a Translation (Conclusion)

- Translation quality is already quite good (and will be improved in the near future)
- Comparable with other converters in this area (however, not comparable with the conversion quality of binary Office to OpenXML conversion tools)
- Working group “Translation” at DIN and Fraunhofer Institute has recently been initiated
 - Analysing the feature translatability between ODF and OpenXML

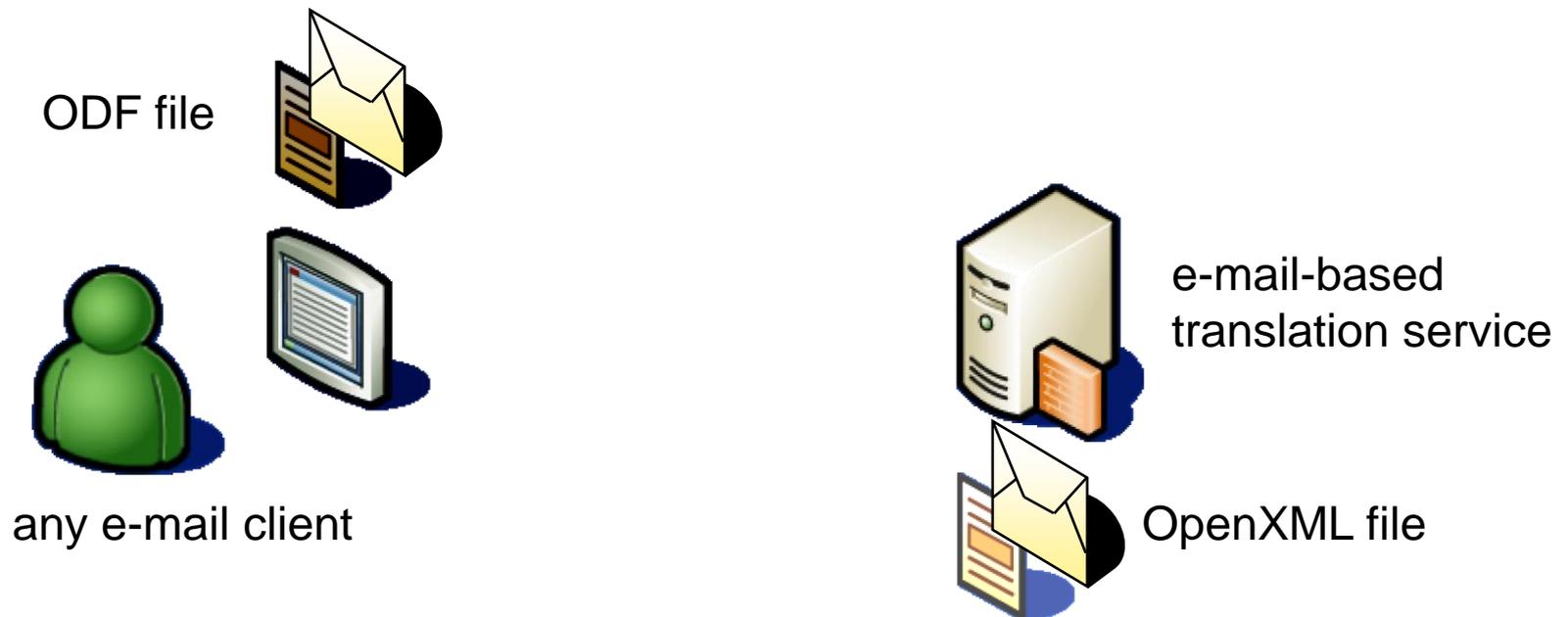
Usage Scenario: Desktop Conversion

- An official in an EU institution downloads an .ODT document from a US website
- The institution uses Word 2007 as its word processing system and has deployed the ODF/OpenXML Translator Add-In for Word 2007 on each workstation
- This add-in enables the official to open, read, revise and save the document again as an .ODT file



Usage Scenario: Central Conversion

- Another EU institution has made a central e-mail-based ODF/OpenXML translation service available to its internal users
- A user in this institution receives an e-mail from a citizen with an .ODT document attached
- The user simply forwards this message to the central conversion service and receives the converted .DOCX document back via e-mail in a matter of minutes



Conclusion

- OpenXML combines the advantages of an open and XML-based format with full compatibility with your existing document store
- The OpenXML/ODF Translator project builds the “interop” bridge to other open standards
- The architecture of the conversion and translation tools allows for various usage scenarios

References

- **OpenXML standard (ECMA)**
 - http://www.ecma-international.org/news/TC45_current_work/
 - http://www.ecma-international.org/news/TC45_current_work/OpenXML%20White%20Paper.pdf
- **File Format Compatibility Pack**
 - <http://www.microsoft.com/downloads/details.aspx?familyid=941b3470-3ae9-4aee-8f43-c6bb74cd1466&displaylang=da>
- **OpenXML/ODF Translator Project**
 - <https://sourceforge.net/projects/odf-converter/>



Thank You!

Q & A

Wolfgang Keber, DIaLOGIKa – Germany, wk@dialogika.de