ODF vs Open XML
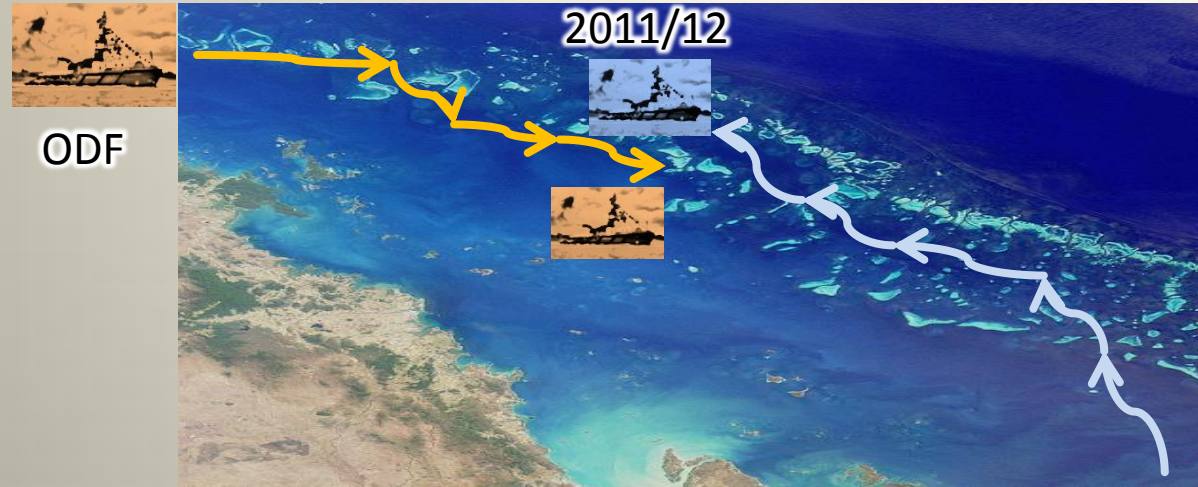# The ISO Specs

Mario Wendt

Chair of the JTC1/SC34 German Mirror & National Standards Officer (NSO) Germany

Microsoft Corporation

2011/12

ODF

uncharted waters

Open XML

DISCOVERING ISO/IEC PDTR 29166 - OPENDOCUMENT FORMAT / OFFICE OPEN XML TRANSLATION - GUIDELINES

Some history to share

- **Establishment NA 043-01-34-01 VT Working Group Translation 29500-26300 by 2007-04-24**

  - The working group Translation 29500 – 26300 (NA 043-01-34-01 VT) which has been set up by NA 043-01-34 AA "Document Description and Processing Languages" working committee has published its initial findings as the first draft of a DIN (German Standards National Body) Technical Report
  - The aim of the "Working Group Translation 29500 - 26300" was to Identify in detail the differences between ODF and Open XML that can help harmonization and interoperability of both file formats in both directions.
  - February 2008 – First Draft of Technical Report at DIN
  - NWIP to JTC1/SC34 approved by 2008-07
  - Start of working on the DIN TR in JTC1/SC34 WG5 October 2008



Initial group

SC34 editors group

Actual Roadmap

- 2010-09-10: at the Tokyo Plenary:
  - Obtain Approval of SC 34 to initiate a PDTR ballot
- 2010-11-30: Start of PDTR ballot
- 2011-02-28: Close PDTR ballot
- 2011-03-31: at Prague meeting, WG 5
  - conducts a **disposition of comments** and
  - obtains SC 34 approval to submit a revised text to **JTC 1** for a DTR ballot
- 2011-05 Start of DTR ballot
- 2011-08: Close DTR ballot
- 2011-09: Disposition of comments on DTR ballot (if any)
- 2011-10: Submittal of a revised text for publication end 2011 or early 2012 Publication as TR

Document structure

- Consideration of official template, distributed after meeting in Okinawa (Sept/Oct 2008)
- Consideration of ODP viewpoints

1. Scope
2. Normative references
3. Terms and definitions
4. Basic principles
5. Use cases
6. Features and functionality
7. Representation and XML structure
8. Translation
9. Examples and tools
10. Conclusion

ODP Viewpoints

- Open Distributed Processing – Reference Model
  - ISO/IEC 10746-1:1998(E)   Overview
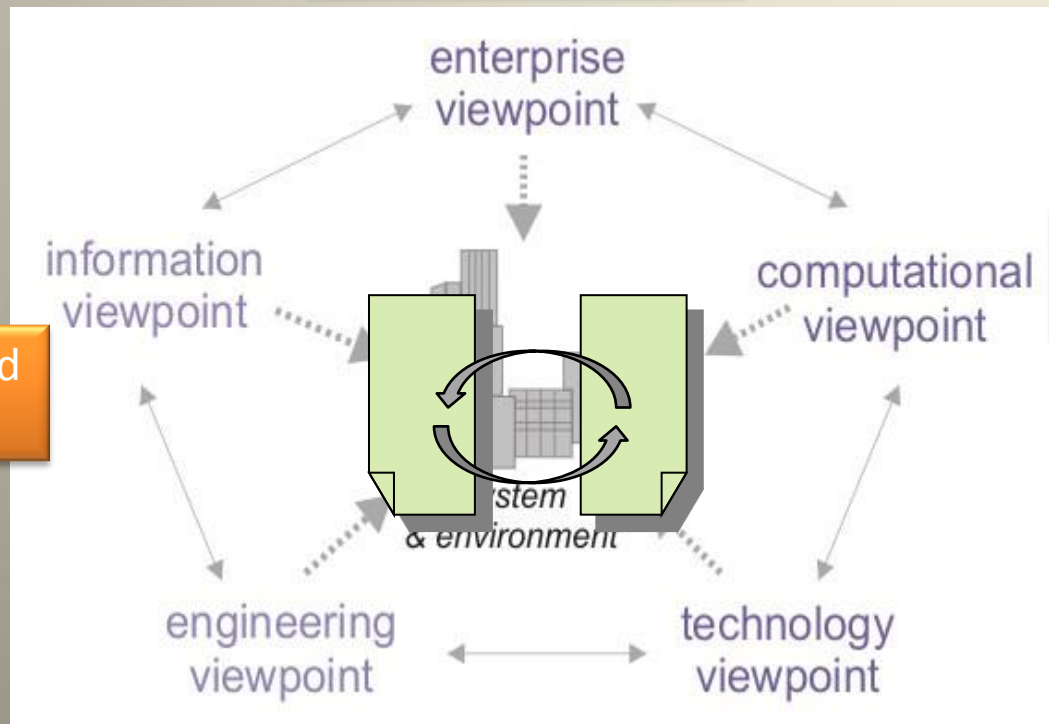 - ISO/IEC 10746-3:1996(E)   Architecture
- Viewpoints
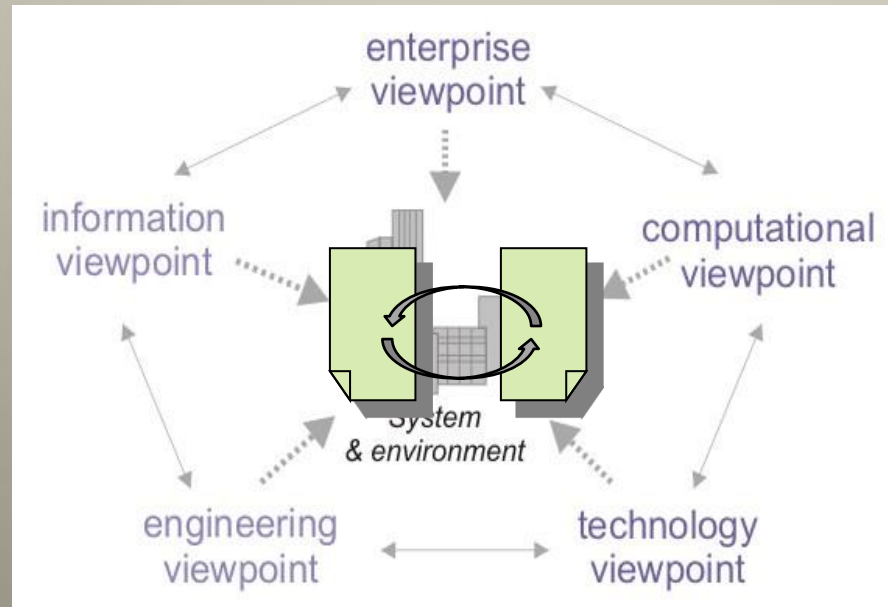
ODP Viewpoints

- In the TR the enterprise view is focusing on **use cases** describing how a document is used in a specific scenario.

- In the TR the computational view is focusing on the **features and functionalities** of a document. The "**what**" is described, independent of "how" the feature is implemented in the particular standard.

- In the TR the information view is focusing on "**how**" the functionality and features of a document are implemented in the standards. The document structure and its **XML markup** are described.

- In the TR the engineering view is focusing on how the features and structures are translated and preserved in the **translation process.**

- In the TR the technical view is focusing on available **resources and tools f**or creating, editing and translating documents.

ODP Viewpoints

Use cases

Features and functionality

Representation and XML structure

enterprise viewpoint

information viewpoint

computational viewpoint

engineering viewpoint

technology viewpoint

stem & environment
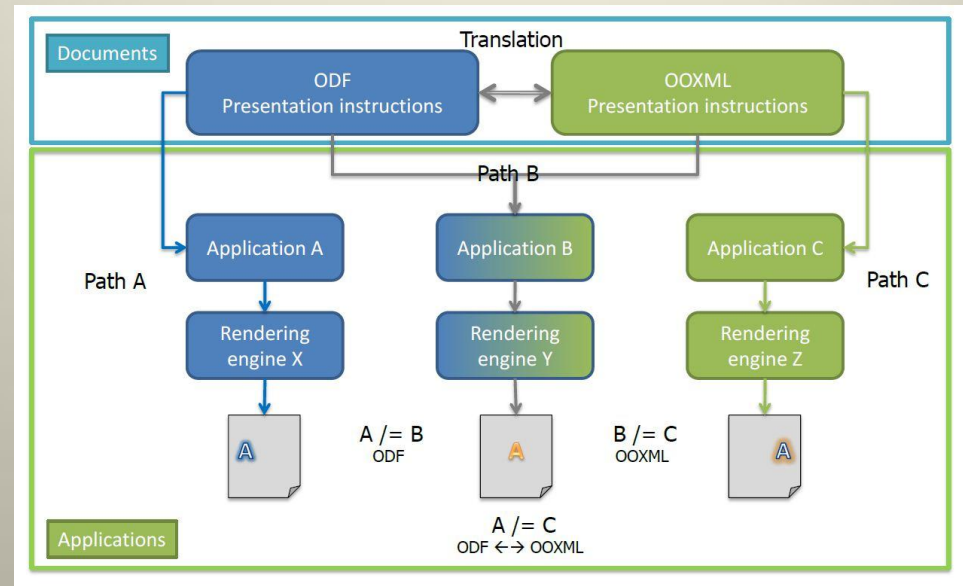
Translation

Examples and tools

Approach

- Define **translation types** and **fidelities**
- Identify document **features** and **functions** from standards
- Use translation types and fidelities to **categorize use cases**
- **Associate** use cases with required features
- Analyze how features and functions can be **translated** between ODF and OOXML

Document properties
Translation fidelities

- **Presentation instructions** include layout and presentation related information such as fonts, spacing, margins, color, and animation in office documents.

- **Document content** covers all properties of content such as text, graphics and formulas defined directly by the author of a document.

- **Dynamic content** covers all aspects of automatically generated content, calculations or form functionalities such as fields, generated tables, or dynamic references.

- **Meta data** cover all information apart from the core document content. Metadata are used to describe meta information about the document such as generator, version, authors, and to ensure the accessibility of documents, for instance by using certificates.

- **Annotation** covers all aspects about annotations to a document, change tracking, and collaborative functions.

- **Document parts** cover all aspects (editing semantics) of structural document properties such as paragraphs, headings, headers, footers, tables, lists, tables, footnotes, indices, and captions.

Document properties
Translation fidelities

Translation types

- One-time translation

  OOXML -> ODF
  ODF -> OOXML

- Round-trip translation

  OOXML -> ODF –> OOXML
  ODF -> OOXML -> ODF

Document properties

- Presentation instructions
- Document content
- Dynamic content
- Meta data
- Annotations
- Document parts

Use Case Template

- **Purpose:**
  - Uniform description of use cases
  - Definition of links between fidelities <-> use case <-> document features

- **Textual description**
  - Describe the **scenario/story** the use case is going to tell
- **Implementation**
  - Describe the **features** that are necessary to implement the use case
- **Requirements/expectation**
  - Describe the **expected behavior** of a translation between the features in both standards
- **Conclusion**
  - Describe in short the **result** of the comparison of the features in both standards

Use case template (cont.)

Use case name: Translation type and properties:	
One-trip translation	✓
Round-trip translation	✓
Presentation instructions	✓
Document content	✓
Dynamic content	✓
Meta data	✓
Annotations	✓
Document parts	✓
Required features: • *Feature a including references to standards* • *Feature b including references to standards*	

Features and functionality

- Word processing documents
  - Text formatting
  - Paragraph formatting
  - Header and footer
  - Tables
  - Itemization and numeration
  - Metadata
  - Indices
  - Change tracking
  - Forms
  - Formulas

- Spreadsheets
  - Formatting
  - Calculation

- Presentations
  - Slides
  - Text formatting
  - Master layouts

- Common features
  - Alternative representations
  - Color models

Use Cases – Word Processing

- Use Cases for word processing documents

  – Empty document

  – Simple text formatting

  – Tables and field functions

  – Itemization and numeration

  – Indices and ToCs

  – Metadata and settings

  – Change tracking and collaborative functions

  – Forms

  – Vector graphic formats

  – Generic fields

  – Font metrics and font substitution

  – Mathematical formulas

- Use Cases for word processing documents (cont.)

  – Font embedding and paper size

  – Chinese language support

  – Korean language support

- Ideas for new use cases
  – Document templates
  – Endnotes and footnotes
  – Glossary documents
  – Subdocuments and books
  – Mail
  – Web settings

Use Cases - Spreadsheets

- Use Cases for spreadsheet documents

  - Empty document

  - Listing and structural features
  - Formulas and calculation

- Ideas for new use cases
  - Access to xml data bases, custom XML docs
  - References across spreadsheets
  - Pivot tables
  - Query tables
  - XML inclusion
  - Real time data
  - Comments

Use Cases - Presentations

- Use Cases for presentation documents

    - Empty document

    - Simple text formatting
    - Itemization and numeration
    - Positioning and layout
    - Slide blending and effects
    - Animations

    - Multimedia content
    - Master layout

- Ideas for new use cases
    - Comments
    - Handouts and notes
    - Slide synchronization

Use Cases – Mutual inclusion / merging of documents

- Use Cases for merged documents

  - Embedded spreadsheet documents
  - Simple text formatting and embedded docs
  - Embedded charts

- Ideas for new use cases
  - Cross references between documents
  - Custom XML data

Empty document

Use case name: Empty document		
Translation type and fidelity		
	One-trip translation	☒
	Round-trip translation	☐
	Graphic fidelity	☒
	Structure fidelity	☐
	Revision fidelity	☐
	Dynamic content fidelity	☐
	Meta data fidelity	☒
	More categories to be added if necessary	☐
	…	

Required features:

- Metadata
- User environment
- Properties and styles
- Document content
- Font information

Simple text formatting

Use case name: Simple text formatting		
Translation type and fidelity		
	One-trip translation	☐
	Round-trip translation ODF -> ooXML -> ODF	☒
	Graphic fidelity	☒
	Structure fidelity	☒
	Revision fidelity	☐
	Dynamic content fidelity	☐
	Meta data fidelity	☐
	More categories to be added if necessary	☐

Required features:

- Text formatting
- Paragraph formatting

Simple text formatting

John Marketer
56 Somerset Lane
Kai Tak, Kowloon
Hong Kong

October 21, 2007

Customer Complaints Centre
GoFast Air
2201 Main Street
London
England

Subject: Complaint on delayed flight Nº PA1234

Dear Sir/Madam,

It is with great regret that I see from delayed flight *Nº PA1234* that GoFast Air is not that reliable as I experienced in the past. The flight took me two hours of waiting on the airport of Hong Kong so that I missed one very important appointment in London. Is there any kind of appropriate compensation that you provide?

Thank you for your prompt attention. I look forward to hearing from you soon.

Best regards,

John Marketer

Tables and field functions

Use case name: Tables and field functions		
Translation type and fidelity		
	One-trip translation ODF ->OOXML	☒
	Round-trip translation	☐
	Graphic fidelity	☒
	Structure fidelity	☒
	Revision fidelity	☐
	Dynamic content fidelity	☐
	Meta data fidelity	☐
	More categories to be added if necessary	☐
	…	
Required features: • Tables		

Tables and field functions

Real Estate Information Bulletin Invoice summary 2007 www.johnmarketer.com/reports		Wednesday, 31 October 2007	
Division & Properties		Results 2007	
		1st Quarter	*2nd Quarter*
Division 1	Rentals	200.000 $	220.000 $
	Sells	430.000 $	440.000 $
Division 2	Rentals	135.000 $	175.000 $
	Sells	390.000 $	376.000 $
Division 3	Rentals	310.000 $	340.000 $
	Sells	612.000 $	685.000 $

Feature description – general template

Feature: featureName

Functionality	Sub-functionality	OOXML	ODF	Translatability	Notes
Textual description	Textual description	Yes/No, Ref to IS 29 500	Yes/No, Ref to IS 26 300	Low/ Medium/ High	Comment

Feature description – example

Feature: Text Formatting

Table 1: Text formatting

Functionality	Sub functionality	OOXML	ODF	Transla-tability	Notes
Bold text (font weight)		Yes 17.3.2.1	Yes 14.6.3	Medium	In addition to bold, ODF allows font weight to be specified numerically (100-900).
Text borders		Yes 17.3.2.4	No	Low	ODF only supports borders on whole paragraphs.
Whitespaces		Yes 17.15.1.18 17.18.7 IS29500-3 10.	Yes 1.6	Medium	Because certain OOXML elements (such as the @preserve attribute defined separately in IS29500 – part 3), are not supported by ODF, translatability of this feature could be problematic.
Capitalization					
	All upper case	Yes 17.3.2.5	Yes 15.4.2	High	
	Small caps	Yes 17.3.2.33	Yes 15.4.1	High	
	All lower case	No	Yes 15.4.2	Low	
Text colour					
	RGB	Yes 17.3.2.6	Yes 14.7.8	High	
	Background colour	Yes 17.3.2.6	Yes 15.4.37	High	
	Based on theme	Yes 17.15.1.20 17.18.97	No	Medium	ODF has no concept of a document theme.
	Blinking text	No	Yes 15.4.36	Low	OOXML supports only blinking backgrounds, but no blinking text.
	Text highlighting	Yes 17.3.2.15	No	Medium	Only a limited range of colours is available for text highlighting.

Conclusions

- Like for many other interoperability related problems it is **impossible** to define a **generic solution** of the translation problem.
  But it seems to be realistic to introduce **subsets or profiles** of document features that are important for specific application areas and that **avoid** fancy features that may be nice to use but prevent interoperability. For such subsets a corresponding document model including mappings to available document formats can be formally defined and validated.
  Such an approach solves the problem resulting from different versions of the document formats and eases the definition of translation rules between the different formats.

- Different tools will in many cases produce different results of a translation between the two document formats. **Therefore the tool must be carefully chosen depending on the given requirements, the available environments and the intention of the document's producers and consumers.**

- A **comprehensive documentation of how** a standard is interpreted and implemented helps a lot to understand the behaviour of the appropriate office suites and the implementation of filters and translation rules.
  Therefore it seems to be desirable to provide comprehensive documentation such as the description of the ODF implementation in Microsoft Office, the community forum of OpenOffice.org, and to further activities like the OASIS interoperability and Conformance TC.
  It should be identified clearly which parts of a standard allow different implementations and probably have to be refined in later versions.

The TR is focusing on the translation between the two standards – not on the features of available implementations.

Thank you for your attention!