

First edition
2011-12-15

**Information technology — Document
description and processing languages —
Guidelines for translation between
ISO/IEC 26300 and ISO/IEC 29500
document formats**

*Technologies de l'information — Description des documents et
langages de traitement — Lignes directrices pour la traduction des
formats de document ISO/CEI 26300 et ISO/CEI 29500*

Reference number
ISO/IEC TR 29166:2011(E)



© ISO/IEC 2011



COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2011

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword	vi
Introduction.....	vii
1 Scope.....	1
2 Normative references.....	1
3 Terms and definitions	2
4 Basic principles	4
4.1 Structure of the report	4
4.1.1 Enterprise view	4
4.1.2 Computational view.....	5
4.1.3 Information view	5
4.1.4 Engineering view	5
4.1.5 Technical view	5
4.2 Approach	6
5 Use cases	8
5.1 Introduction.....	8
5.2 Word processing documents	8
5.2.1 Empty document	8
5.2.2 Simple text and paragraph formatting	10
5.2.3 Asian language support	11
5.2.4 Line breaks in East Asian text	14
5.2.5 Text direction	16
5.2.6 Phonetic guide functions	18
5.2.7 Tables and field functions	20
5.2.8 Footnotes and endnotes.....	22
5.2.9 Itemization and numeration.....	24
5.2.10 Indices and tables of contents.....	25
5.2.11 Metadata and settings.....	26
5.2.12 Change tracking and collaboration support.....	28
5.2.13 Bibliographies and optional document parts	31
5.2.14 Sub documents and books	32
5.2.15 Forms.....	35
5.2.16 Vector graphics	36
5.2.17 Font embedding and paper size	38
5.2.18 Font metrics and font substitution	40
5.2.19 Document fields.....	41
5.2.20 Inclusion of user defined XML	43
5.2.21 Mathematical formulas.....	45
5.3 Spreadsheet documents.....	46
5.3.1 Empty spreadsheet document	46
5.3.2 Listing and structural features.....	48
5.3.3 Formulas and calculation	49
5.4 Presentation documents.....	51
5.4.1 Empty presentation document.....	51
5.4.2 Simple text formatting.....	52
5.4.3 Itemization and numeration.....	54
5.4.4 Positioning and layout.....	55
5.4.5 Slide blending and animation effects	57
5.4.6 Animations	58
5.4.7 Comments	60
5.4.8 Multimedia content.....	62

5.4.9	Master layout.....	64
5.5	Common properties and mutual inclusion of documents.....	65
5.5.1	Hyperlinks between documents.....	65
5.5.2	Colours	67
5.5.3	Embedded spreadsheet documents.....	69
5.5.4	Simple text formatting and embedded documents.....	71
5.5.5	Embedded charts.....	73
6	Features and functionality	74
6.1	Introduction	74
6.2	Word processing documents	75
6.2.1	Text formatting.....	75
6.2.2	Paragraph formatting	77
6.2.3	Header and footer	82
6.2.4	Tables.....	82
6.2.5	Itemization and numeration	84
6.2.6	Metadata language entries.....	86
6.2.7	Indices.....	86
6.2.8	Change tracking and collaborative functions.....	87
6.2.9	Bibliographies and optional document parts	88
6.3	Spreadsheet documents	89
6.3.1	Introduction	89
6.3.2	Formatting	89
6.3.3	Calculation.....	90
6.3.4	Additional properties.....	94
6.4	Presentation documents.....	95
6.4.1	Introduction	95
6.4.2	Slides	95
6.4.3	Text formatting.....	96
6.4.4	Master layout.....	97
6.5	Common aspects.....	98
6.5.1	Alternative presentations.....	98
6.5.2	Colour models.....	99
6.5.3	Custom XML parts	100
6.5.4	Packages	100
7	Representation and XML structure.....	102
7.1	Introduction	102
7.2	Word processing documents	103
7.2.1	Logical structure.....	103
7.2.2	Paragraphs	105
7.2.3	Styles	106
7.2.4	Tables.....	107
7.2.5	Lists - Itemization and numeration	111
7.2.6	Indices.....	112
7.2.7	Change tracking and collaboration support	115
7.2.8	Section and page layout	118
7.3	Spreadsheet documents	120
7.3.1	Logical structure.....	120
7.3.2	Table contents.....	121
7.3.3	Table style	123
7.3.4	Formulas and calculation	124
7.3.5	Charts.....	126
7.4	Presentation documents.....	128
7.4.1	Logical structure.....	128
7.4.2	Text formatting.....	129
7.4.3	Master layout.....	131
7.4.4	Animations	132
7.5	Summary.....	137
8	Translation.....	137

8.1	Introduction.....	137
8.2	Translation complexity	137
8.3	Sample translations	139
8.3.1	Easy translation.....	139
8.3.2	Moderate translation	143
8.3.3	Difficult translations.....	149
8.4	Guidelines for evaluating translatability.....	150
8.4.1	Translation fidelity.....	151
8.4.2	Document interoperability.....	152
9	Examples and tools.....	153
10	Conclusion	154
10.1	Resume.....	155
	Bibliography.....	156

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

In exceptional circumstances, when the joint technical committee has collected data of a different kind from that which is normally published as an International Standard ("state of the art", for example), it may decide to publish a Technical Report. A Technical Report is entirely informative in nature and shall be subject to review every five years in the same manner as an International Standard.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

ISO/IEC TR 29166 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 34, *Document description and processing languages*.

Introduction

OASIS Open Document Format ODF 1.0 (ISO/IEC 26300) and Office Open XML (ISO/IEC 29500) are both open document formats for saving and exchanging word processing documents, spreadsheets and presentations. Both formats are XML based but differ in design and scope.

OASIS ODF 1.0 was published by OASIS as an *OASIS standard* in May 2005. The second edition of ODF 1.0 has been published by OASIS as a *committee specification* in July 2006 and accepted as an International Standard by ISO (ISO/IEC 26300) in December 2006. Office Open XML was first approved in December 2006 by the ECMA International General Assembly as ECMA-376. An updated version was published in November 2008 by ISO (ISO/IEC 29500). The corresponding version, ECMA-376 2nd edition, was published in December 2008.

This Technical Report addresses both developers seeking to implement either the OpenDocument or the Office Open XML International Standard and template designers and other power users whose competences cut across the spectrum of XML and XML-related technologies which directly or remotely deal with one or both of the two International Standards. This Technical Report will be of great assistance to those seeking to exchange documents between formats, to extract data from or import data into documents, or to write applications supporting the two formats.

This Technical Report aims at analysing the two International Standards and their underlying concepts in terms of interoperability issues for a selected set of features. It analyses the way these features are implemented in both International Standards and estimates the degree of translatability between them using a table-based comparison. This Technical Report serves as a preliminary technical translation guideline for evaluating translatability between certain parts of the two International Standards. It does not compare different implementations which can cause additional kinds of interoperability problems.

Both Office Open XML and OpenDocument formats are basically descriptions of schemas used for word processing documents, spreadsheets and presentations created by office application suites. Both are open formats. A key design objective is to guarantee long-term access to data without the legal or technical barriers associated with proprietary binary formats. XML schema definitions are normative parts of both International Standards.

Manipulating documents is fundamentally facilitated by separating a document's layout from its content. Editing the layout and data components independently of one another affords considerable flexibility in creating and editing office documents. Defining the structure and content of documents has been the focus of both International Standards. A document's layout is ultimately governed by the implementation of the office suite, in particular by the rendering engine. Thus, as depicted in Figure 1, using exactly the same standard to describe a document does not guarantee that different office suites will produce identical layouts. Consequently, this Technical Report focuses more on the definition of guidelines for the translation of document structure, content and presentation instructions than on the preservation of document layout.

In this Technical Report the two International Standards will be examined in their universality and not by comparing specific implementations such as Microsoft Office or OpenOffice.org/LibreOffice. For this reason, various examples have been developed using a simple XML editor which supports both standards. The names of specific implementations may be used in the use cases to illustrate the real world scenario behind the use case. The figures in this Technical Report are created for illustration purposes, using available tools such as OpenOffice 3.* and Microsoft Office 2010. It should not be assumed that the current versions of these implementations support all the features needed to implement the use case, especially the document standards and the translation between them.

Several use cases do not mention existing tools, but rather use abstract names such as document format A (DF-A) and document format B (DF-B).

This Technical Report begins with a presentation of typical use cases characterizing scenarios where specific features supported by both document formats are used. It then analyses the most important features of one

document format and shows how those features can best be represented in the other format. It then reviews the concepts and various features of the two document formats in order to provide a good understanding of the formats' common features and especially their differences. Most features can be translated to the other format with varying degrees of fidelity. For the most important features, this Technical Report provides detailed information on the implementation of the feature and the extents to which that feature can be translated, including typical translation rules. Finally, an overview summing up the most important results and deriving guidelines for the translation between both formats concludes this Technical Report.

The following abbreviations are used throughout this Technical Report:

- *ODF*, which stands for OpenDocument Format (ISO/IEC 26300:2006);
- *OOXML*, which stands for Office Open XML (ISO/IEC 29500:2008).

It is hoped that this Technical Report will be useful in understanding how the ODF and OOXML International Standards compare and how their functionality can be mapped between the two formats. It is a necessary step to the goal of helping achieve interoperability and harmonization between the two formats.

History of ODF and OOXML

ODF was originally developed by Sun Microsystems between 2000 and 2002 with the following objective:

*“To create as a community, the leading international office suite that will run on all major platforms and provide access to all functionality and data through open-component based APIs and an XML-based file format.”*¹

In 2002, the standardization process was initiated at OASIS, and in 2005 the standard was published as *OASIS Open Document Format for Office Applications*, abbreviated as *OpenDocument* or *ODF*. In 2006, *Open Document Format for Office Applications v.1.0* became an ISO International Standard (ISO/IEC 26300). *Open Document Format for Office Applications v.1.1* has been published by OASIS as an *OASIS standard* in February 2007. At the time of writing (June 2011) Version 1.2 has been released as a Committee Specification 1.0. While version 1.0 of the ODF standard only consists of one part, the current version is structured into three parts: core, formulas, and packages.

Microsoft followed suit in 2006 via the *Open Specification Promise (OSP²)* by opening the format of its 2007 version of the Microsoft office suite (version 12) for which it also uses XML as an exchange and storage format. OOXML is a file format originally developed by Microsoft as a successor to its earlier Office 2003 file formats. It is used for representing spreadsheet, presentation and word processing documents. In 2006 Office Open XML became an ECMA standard (ECMA-376, 1st edition). In 2008, a revised version of ECMA-376 became an ISO International Standard (ISO/IEC 29500:2008), which has its equivalent in the ECMA-376, 2nd edition.

ISO/IEC 29500 is structured into four parts, each of which contains normative as well as informative material: Fundamentals and Markup Language Reference, Open Packaging Conventions, Markup Compatibility and Extensibility, and Transitional Migration Features.

At the time of writing (June 2011) the following corrigenda and amendments have been published:

- ISO/IEC 29500-1:2008/Cor.1:2010, ISO/IEC 29500-2:2008/Cor.1:2010, ISO/IEC 29500-3:2008/Cor.1:2010 and ISO/IEC 29500-4:2008/Cor.1:2010, containing minor technical corrections and editorial modifications;
- ISO/IEC 29500-1:2008/Amd.1:2010 and ISO/IEC 29500-4:2008/Amd.1:2010, containing namespace changes and modifications concerning the usage of percentage (%) values;

¹ http://www.openoffice.org/about_us/ooo_release.html

² <http://www.microsoft.com/interop/osp/default.mspx>

- ISO/IEC 29500:2011 (ECMA 376 3rd edition) as a consolidated version of OOXML containing the above-mentioned corrigenda and amendments;
- ISO/IEC 26300:2006/Cor.1:2010, containing editorial modifications;
- ISO/IEC 26300:2006/Cor.2:2011, fixing editorial errors.

In addition, the following Amendments are under preparation:

- Amendment 1 to ISO/IEC 29500-1:2011 and Amendment 1 to ISO/IEC 29500-4:2011 about ISO 8601 dates;
- Amendment 1 to ISO/IEC 26300:2006 introducing ODF 1.1.

Information technology — Document description and processing languages — Guidelines for translation between ISO/IEC 26300 and ISO/IEC 29500 document formats

1 Scope

This Technical Report provides guidelines for translation between ISO/IEC 26300 and ISO/IEC 29500 document formats. It starts by studying common use cases to identify how the most important functionalities of one document format can be represented in the other format. This is followed by a thorough review of the concepts, architectures and various features of the two document formats in order to provide a good understanding of the commonalities and differences. It is expected that functionalities will be able to be translated with different degrees of fidelity to the other format. As an illustrative sample of this functionality, detailed information is provided on the extent to which those functionalities can be translated. This Technical Report is a necessary step to the goal of helping achieve interoperability and harmonization between the two formats.

2 Normative references

The following referenced documents are indispensable for the application of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 26300:2006, *Information technology — Open Document Format for Office Applications (OpenDocument) v1.0*

ISO/IEC 29500-1:2008, *Information technology — Document description and processing languages — Office Open XML File Formats — Part 1: Fundamentals and Markup Language Reference*

ISO/IEC 29500-2:2008, *Information technology — Document description and processing languages — Office Open XML File Formats — Part 2: Open Packaging Conventions*

ISO/IEC 29500-3:2008, *Information technology — Document description and processing languages — Office Open XML File Formats — Part 3: Markup Compatibility and Extensibility*

ISO/IEC 29500-4:2008, *Information technology — Document description and processing languages — Office Open XML File Formats — Part 4: Transitional Migration Features*