TECHNICAL REPORT

ISO/IEC TR 29166

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 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

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* (OSP[2]) 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*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

3.1
translation type
methods used when translating between ODF and OOXML documents

NOTE This Technical Report distinguishes four *translation types*:

- one way ODF to OOXML translation;
- one way OOXML to ODF translation;
- round trip ODF to OOXML to ODF translation;
- round trip OOXML to ODF to OOXML translation.

3.2
translation fidelity
quality of a translation process between the ODF and OOXML *document formats*

NOTE 1 *Translation fidelity* depends on *document properties*.

NOTE 2 *Translation fidelity* cannot be measured in an absolute manner; it depends on the intentions of the document's authors.

3.3
document type
characterization of the specific purpose and content of a document

NOTE 1 This Technical Report distinguishes three major *document types*: word processing, spreadsheet and presentation documents.

NOTE 2 Some *document features* only exist in one *document type*; other *features* have been defined for more than one *document type*.

NOTE 3 The association between *document type* and *document feature* can be different for ODF and OOXML.

3.4
document property
description of different yet independent dimensions within the specification of a document

NOTE 1 As defined in 4.2 this Technical Report distinguishes the following *document properties*:

- presentation instructions;
- content;
- dynamic content;
- meta data;
- annotations and security;
- document parts.

NOTE 2 *Document properties* are implemented using *document features*.

3.5
document feature
characterization of a document used to implement specific aspects of a *document property*

NOTE 1 *Document features* are visible to a user.

NOTE 2 *Document features* are illustrated by associated use cases in this Technical Report.

NOTE 3 The terms *feature*, *functionality* and *sub functionality* are used to structure the comparison of both *document formats* in Clause 6.

NOTE 4 *Document features* implement *document properties*.

3.6
document format
synonym for *document standard* within this Technical Report

3.7
functionality
refinement of *document features*

NOTE 1 For example, the format of a paragraph is a *feature* and the height of a line is a *functionality*.

NOTE 2 In many cases *functionality* is implemented using XML types or elements.

3.8
sub functionality

itemization of *functionality*

NOTE 1 For example, the height of a line can be defined as fixed, font-independent, automatic, etc.

NOTE 2 In many cases *sub functionality* is implemented using XML attributes.

3.9
translatability level
rough scale for translation fidelity

NOTE 1 Translatability *levels* are used in Clause 6.

NOTE 2 *Translatability levels* have a three tier range (low, medium, high).

3.10
translation complexity
description of the complexity of the translation process for *document features*, considering their structures and associated translation rules

NOTE 1 *Translation complexity* is a three value metric system (easy, moderate, difficult).

NOTE 2 *Translation complexity* is used in Clause 8 to classify the translation of *document features* or *functionalities* from one format to the other.

4 Basic principles

4.1 Structure of the report

The report is structured according to the viewpoints introduced in the reference model for Open Distributed Processing ODP (ISO/IEC 10746). Refer to ISO/IEC 10746-1:1998 and ISO/IEC 10746-3:1996.

4.1.1 Enterprise view

The enterprise viewpoint is concerned with the purpose, scope and policies governing the activities of the specified system within the organization of which it is a part. All requirements that are relevant to defining the architecture and properties of the system are gathered in this viewpoint.

In clause 5 the TR describes the translation process from the enterprise viewpoint. It focuses on use cases that describe how a document is used in a specific scenario. Features and functionalities of documents like presentation instructions, structural information, application context, and the content itself as well as certain conformance classes based on translation types and properties have been taken into consideration. Users of document standards and decision makers are the intended readers of this section.

4.1.1.1 Use case template

To facilitate comparisons and a quick overview, use cases are described using the following template:

Textual description:

- *Describe the scenario/story the use case is going to tell;*
- *Include one or more figures demonstrating the use case (optional);*
- *Define the translation type and fidelity to be demonstrated.*

Implementation:

- *Describe the features necessary to implement the use case.*

Use case[3] name: **Translation type and properties:**	
One-trip translation	ODF → OOXML or/and OOXML → ODF
Round trip translation	ODF→OOXML→ODF or/and OOXML→ODF→OOXML
Presentation instructions	✓
Document content	✓
Dynamic content	✓
Metadata	✓
Annotations and security	✓
Document parts	✓

[3] For details about the use case table please refer to section 4.2.

Additional properties	...
Required features: • *Feature a including references to standards* • *Feature b including references to standards*	

Requirements:

- *Describe the expected behaviour of a feature's translation between both International Standards;*
- *Describe how the document(s) used in the use case should be defined to achieve the intended fidelity.*

Conclusion:

- *Compare the applicable features in both International Standards and the translation rules and fidelity as elaborated on in clauses* 6 *and* 8*.*

4.1.2 Computational view

The computational viewpoint is concerned with the functional decomposition of a system into a set of objects that interact at interfaces: thus enabling system distribution.

In clause 6 the TR describes the translation process from the computational viewpoint. It focuses on the features and functionalities of a document. The *what* is described, independent of *how* the feature is implemented in the particular standard. Power users and developers are the intended readers of this section.

4.1.3 Information view

The information viewpoint is concerned with the kinds of information handled by a system and constraints on the use and interpretation of that information. An information specification of a system defines the semantics of information and the semantics of information processing.

In clause 7 the TR describes the translation process from the information viewpoint. It 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. Power users, developers and persons responsible for the maintenance of the standards are the intended readers of this section.

4.1.4 Engineering view

The engineering viewpoint is concerned with the infrastructure required to support system distribution. An engineering specification defines the mechanisms and functions required to support distributed interaction between objects.

In clause 8 the TR describes the translation process from the engineering viewpoint. It focuses on how the features and structures are translated and preserved in the translation process. Developers and persons responsible for the maintenance of the International Standards are the intended readers of this section.

4.1.5 Technical view

The technology viewpoint is concerned with the choice of technology used to support system distribution.