INTERNATIONAL STANDARD

ISO/IEC 29500-1

Second edition
2011-08-15

Information technology — Document description and processing languages — Office Open XML File Formats —

Part 1:
Fundamentals and Markup Language Reference

Technologies de l'information — Description des documents et langages de traitement — Formats de fichier "Office Open XML" —

Partie 1: Principes essentiels et référence de langage de balisage

Table of Contents

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.

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 29500-1 was prepared by ISO/IEC JTC 1, Information technology, Subcommittee SC 34, Document description and processing languages.

This second edition cancels and replaces the first edition (ISO/IEC 29500-1:2008), which has been technically revised by incorporation of the Amendment ISO/IEC 29500-1:2008/Amd.1:2010 and the Technical Corrigendum ISO/IEC 29500-1:2008/Cor.1:2010.

ISO/IEC 29500 consists of the following parts, under the general title *Information technology — Document description and processing languages — Office Open XML File Formats*:

- *Part 1: Fundamentals and Markup Language Reference*
- *Part 2: Open Packaging Conventions*
- *Part 3: Markup Compatibility and Extensibility*
- *Part 4: Transitional Migration Features*

Annexes A, 7 and 8 form a normative part of this Part of ISO/IEC 29500. Annexes B–- and =–U are for information only.

This Part of ISO/IEC 29500 includes five annexes (Annex A, Annex B, Annex F, Annex G, and Annex H) that refer to data files provided in electronic form.

Introduction

ISO/IEC 29500 specifies a family of XML schemas, collectively called *Office Open XML*, which define the XML vocabularies for word-processing, spreadsheet, and presentation documents, as well as the packaging of documents that conform to these schemas.

The goal is to enable the implementation of the Office Open XML formats by the widest set of tools and platforms, fostering interoperability across office productivity applications and line-of-business systems, as well as to support and strengthen document archival and preservation, all in a way that is fully compatible with the existing corpus of Microsoft Office documents.

The following organizations have participated in the creation of ISO/IEC 29500 and their contributions are gratefully acknowledged:

Apple, Barclays Capital, BP, The British Library, Essilor, Intel, Microsoft, NextPage, Novell, Statoil, Toshiba, and the United States Library of Congress

Information technology — Document description and processing languages — Office Open XML File Formats

Part 1:
Fundamentals and Markup Language Reference

1. Scope

ISO/IEC 29500 defines a set of XML vocabularies for representing word-processing documents, spreadsheets and presentations. On the one hand, the goal of ISO/IEC 29500 is to be capable of faithfully representing the pre-existing corpus of word-processing documents, spreadsheets and presentations that had been produced by the Microsoft Office applications (from Microsoft Office 97 to Microsoft Office 2008, inclusive) at the date of the creation of ISO/IEC 29500. It also specifies requirements for Office Open XML consumers and producers. On the other hand, the goal is to facilitate extensibility and interoperability by enabling implementations by multiple vendors and on multiple platforms.

This Part of ISO/IEC 29500 specifies concepts for documents and applications of both strict and transitional conformance.

2. Conformance

The text in ISO/IEC 29500 is divided into *normative* and *informative* categories. Text marked informative (using the mechanisms described in §7) is for information purposes only. Unless stated otherwise, all text is normative.

Use of the word "shall" indicates required behavior.

Any behavior that is not explicitly specified by ISO/IEC 29500 is implicitly unspecified (§4).

2.1   Goal

This subclause is informative

The goal of this clause is to define conformance, and to provide interoperability guidelines in a way that fosters broad and innovative use of the Office Open XML file format, while maximizing interoperability and preserving investment in existing files and applications (§4). By meeting this goal, ISO/IEC 29500 benefits the following audiences:

- Developers that design, implement, or maintain Office Open XML applications.
- Developers that interact programmatically with Office Open XML applications.
- Governmental or commercial entities that procure Office Open XML applications.
- Testing organizations that verify conformance of specific Office Open XML applications to ISO/IEC 29500. (Note that ISO/IEC 29500 does not include a test suite.)
- Educators and authors who teach about Office Open XML applications.

End informative subclause

2.2   Issues

This subclause is informative

To achieve the above goal, the following issues need to be considered:

1. The application domain encompasses a range of possible consumers (§4) and producers (§4) so broad that defining specific application behaviors would restrict innovation. For example, stipulating visual layout would be inappropriate for a consumer that extracts data for machine consumption, or that renders text in sound. Another example is that restricting capacity or precision runs the risk of diluting the value of future advances in hardware.
2. Commonsense user expectations regarding the interpretation of an Office Open XML package (§4) play such an important role in that package's value that a purely syntactic definition of conformance would fail to effect a useful level of interoperability. For example, such a definition would admit an application that reads a package, and then writes it in a manner that, though syntactically valid, differs arbitrarily from the original.

3. Legitimate operations on a package include deliberate transformations, making blanket change prohibitions inappropriate in the conformance definition. For example, collapsing spreadsheet formulas to their calculated values, or converting complex presentation graphics to static bitmaps, could be correct for an application whose published purpose is to perform those operations. Again, commonsense user expectation makes the difference.

4. Existing files and applications exercise a broad range of formats and functionality that, if required by the conformance definition, would add an impractical amount of bulk to ISO/IEC 29500 and could inadvertently obligate new applications to implement a prohibitive amount of functionality. This issue is caused by the breadth of currently available functionality and is compounded by the existence of legacy formats.

End informative subclause

2.3   What ISO/IEC 29500 Specifies

To address the issues listed above, ISO/IEC 29500 constrains both syntax and semantics, but it is not intended to predefine application behavior. Therefore, it includes, among others, the following three types of information:

1. W3C XML Schemas and an associated validation procedure for validating document syntax against those schemas. (The validation procedure includes un-zipping, locating files, processing the extensibility XML elements and attributes, and W3C XML Schema validation.)

2. Additional syntax constraints in written form. [*Note*: These constraints are described in written form because they could not feasibly be expressed in the schema language. *end note*]

3. Descriptions of XML element semantics. The semantics of an XML element refers to its intended interpretation by a human being.

2.4   Document Conformance

Document conformance is purely syntactic; it involves only Items 1 and 2 in §2.3 above.

- A conforming document shall conform to the transitional W3C XML Schema or the strict W3C XML Schema (Item 1), and any additional syntax constraints (Item 2).
- The document shall be of category Wordprocessing, Spreadsheet, or Presentation.
- The document character set shall conform to the Unicode Standard and ISO/IEC 10646, with either the UTF-8 or UTF-16 encoding form, as required by the XML 1.0 standard.
- Any XML element or attribute not explicitly included in ISO/IEC 29500 shall use the extensibility mechanisms described by ISO/IEC 29500-1 and ISO/IEC 29500-3.

Each Part of this multi-part standard has its own conformance clause. The term *conformance class* is used to disambiguate conformance within different Parts of this multi-part standard. This Part of ISO/IEC 29500 defines the following document conformance classes:

- *WML Strict*, if the document is a conforming document of category Wordprocessing that conforms to the strict schema and does not include any features from Part 4.

- *SML Strict*, if the document is a conforming document of category Spreadsheet that conforms to the strict schema and does not include any features from Part 4.
- *PML Strict*, if the document is a conforming document of category Presentation that conforms to the strict schema and does not include any features from Part 4.

In addition, documents of conformance class WML Strict, SML Strict, or PML Strict shall not embed documents of conformance class WML Transitional, SML Transitional, or PML Transitional as defined in Part 4.

Document categories Wordprocessing, Spreadsheet, and Presentation are defined in §4.

[*Note*: Other document conformance classes could be defined in the future. *end note*]

[*Note*: A document cannot be of more than one of the above conformance classes. *end note*]

2.5   Application Conformance

Application conformance incorporates both syntax and semantics; it involves items 1, 2, and 3 in §2.3 above.

- A conforming consumer shall not reject any conforming documents of at least one document conformance class.
- A conforming producer shall be able to produce conforming documents of at least one document conformance class.
- A conforming application shall treat the information in Office Open XML documents in a manner consistent with the semantic definitions given in ISO/IEC 29500. An application's intended behavior need not require that application to process all of the information in an Office Open XML document. However, the information that it does process shall be processed in a manner that is consistent with the semantic definitions given in ISO/IEC 29500.

[*Note*: This note illustrates the third bullet above. Conforming applications might serve various functions. Examples include a viewer, an editor, and a back-end processor. Here is an illustration of how the third bullet applies to each of those examples:

- If a conforming viewer supports a given feature, then when it displays information using that feature, it respects the semantics of that feature as described in the Standard.
- If a conforming editor supports a given feature, then when it provides its user with an interface for manipulating information using that feature, it respects the semantics of that feature as described in the Standard.
- If a conforming back-end processor supports a given feature, then when that processor transforms or assembles information involving that feature, that processor respects the semantics of that feature as described in the Standard.

end note]

This Part of ISO/IEC 29500 defines the following application conformance classes:

- *WML Strict*, if the application is a conforming application that is a consumer or producer of documents having conformance class WML Strict.
- *SML Strict*, if the application is a conforming application that is a consumer or producer of documents having conformance class SML Strict.
- *PML Strict*, if the application is a conforming application that is a consumer or producer of documents having conformance class PML Strict.

Conformance can also involve the use of application descriptions; see §2.6 for details.

2.6   Application Descriptions

An application can be defined as conforming to zero or more *application descriptions* in a particular conformance class.

The application descriptions defined within ISO/IEC 29500 are:

- Base
- Full

[*Note*: These application descriptions should not be taken as limiting the ability of an application provider to create innovative applications. They are intended as a mechanism for labelling applications rather than for restricting their capabilities. The intention is to promote interoperability between different applications that share the same conformance class. Application descriptions are orthogonal to the conformance of the documents produced by those applications. For example, a tool used for automated translation of documents might have an application description of "Base" but will still produce fully conformant documents. *end note*]

The application descriptions are determined in terms of an application's semantic understanding of particular features. *Semantic understanding* is to be interpreted in that an application shall treat the information in Office Open XML documents in a manner consistent with the semantic definitions given in ISO/IEC 29500.

Each application description is identified by a URI.

The application descriptions are defined in the following subclauses.

2.6.1   Base Application Description

Description URI: http://purl.oclc.org/ooxml/descriptions/base

An application conforming to this description has a semantic understanding of at least one feature within its conformance class.

[*Note*: In addition, applications that include a user interface are strongly recommended to support all accessibility features appropriate to that user interface. *end note*]

2.6.2   Full Application Description

Description URI: http://purl.oclc.org/ooxml/descriptions/full

An application conforming to this description has a semantic understanding of every feature within its conformance class.

2.6.3 Additional Application Descriptions

It is expected that additional application descriptions will be defined within the maintenance process for ISO/IEC 29500. It is also expected that third parties might define their own application descriptions; for example to inform their procurement decisions, or to deal with domains such as accessibility.

[*Note*: A possible application description would be a "standard" application description for a wordprocessing application. This could be created by taking the intersection of the features available in common wordprocessing applications such as Word 2000, OpenOffice 2, WordPerfect, and iWork Pages. In addition, it could define formats such as specific image and video formats required to be supported to conform to the description. Similar descriptions could be created for spreadsheet applications and presentation applications. Such a description would promote interoperability between applications implementing OOXML. It would also promote interoperability between applications implementing OOXML and applications implementing other document formats such as ISO/IEC 26300. *end note*]

Application descriptions are not required to be strict subsets of each other. An application can simultaneously conform to multiple application descriptions.

Any such newly created description shall enumerate the features that are required for conformance to it. Such a description should provide a machine-processable schema, preferably using a standard such as ISO/IEC 19757.

[*Note*: If the application conforming to a description is a document consumer, it should be able to consume any document that respects such a schema associated with the description. If the application is a document producer, any document produced by that application should respect the schema of the description. *end note*]

Any such description should be identified using a URI, in a similar manner to the names used for application descriptions within ISO/IEC 29500.

[*Note*: For the convenience of users of the description, it is recommended that creators of a description should make a human- or machine-readable form of that description available at a URL corresponding to the description URI. *end note*]

2.6.4 Representation of Application Descriptions within Documents

An application description is related to applications, rather than to document conformance. Therefore, there is no normative mechanism for representing an application description within a document.

[*Note*: It is recommended that implementers wishing to represent an application description within a document use the standard metadata mechanism for Office Open XML. *end note*]

2.7 Interoperability Guidelines

[*Guidance*: The following interoperability guidelines incorporate semantics (Item 3 in §2.3 above).

For the guidelines to be meaningful, a software application should be accompanied by documentation that describes what subset of ISO/IEC 29500 it supports. The documentation should highlight any behaviors that would, without that documentation, appear to violate the semantics of document XML elements. Together, the application and documentation should satisfy the following conditions.

1. The application need not implement operations on all XML elements defined in ISO/IEC 29500. However, if it does implement an operation on a given XML element, then that operation should use semantics for that XML element that are consistent with ISO/IEC 29500.
2. If the application moves, adds, modifies, or removes XML element instances with the effect of altering document semantics, it should declare the behavior in its documentation.

The following scenarios illustrate these guidelines.

- A presentation editor that interprets the preset shape geometry "`rect`" as an ellipse does not observe the first guideline because it implements "`rect`" but with incorrect semantics.
- A batch spreadsheet processor that saves only computed values even if the originally consumed cells contain formulas, might satisfy the first condition, but does not observe the second because the editability of the formulas is part of the cells' semantics. To observe the second guideline, its documentation should describe the behavior.
- A batch tool that reads a word-processing document and reverses the order of text characters in every paragraph with "Title" style before saving it can be conforming even though ISO/IEC 29500 does not recommend this behavior. This tool's behavior would be to transform the title "Office Open XML" into "LMX nepO eciffO". Its documentation should declare its effect on such paragraphs.

The normative requirements in §2.4 imply that a conforming producer shall not write unescaped non-XML characters. As an implementation guideline, a conforming producer additionally should not write escaped non-XML characters. Doing so damages interoperability with existing XML-based standards such as SOAP and RDF. For example, implementers could either refuse to create documents including such characters, or warn users that including such characters compromises the re-usability of their documents. *end guidance*]

3. 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.

ANSI X3.4-1986, *American Standard Code for Information Interchange (ASCII)*

Bureau of Standards, Metrology and Inspection of the Ministry of Economic Affairs, *CNS 7648: Data Elements and Interchange Formats — Information Interchange — Representation of Dates and Times*

Calendar Reform Committee, Indian Ephemeris and Nautical Almanac. 1957

Stokes, M., M. Anderson, S. Chandrasekar, and R. Motta. *A Standard Default color Space for the Internet. Vers. 1.10*. November 5, 1996. http://www.w3.org/Graphics/Color/sRGB

Har'El, Zvi, *Gauss Formula for the Julian Date of Passover*. Deptartment of Mathematics, Technion, Israel Institue of Technology, Haifa 32000, Isreal, 2005, 6

Duerst, M, and M Suignard. *Internationalized Resource Identifiers (IRIs).* IETF. January 2005. http://tools.ietf.org/html/rfc3987

IANA, *Character Sets from IANA*, as specified at http://www.iana.org/assignments/character-sets

IANA. *MIME Media Types*. Internet Assigned Numbers Authority. http://www.iana.org/assignments/media-types/

IEC 60559:1989, *Binary Floating-Point Arithmetic for Microprocessor Systems*

ISO/IEC 2382-1:1993, *Information technology — Vocabulary — Part 1: Fundamental terms*

ISO 8601:2004, *Data elements and interchange formats — Information interchange — Representation of dates and times*

ISO/IEC 8859-1:1998, *Information technology — 8-bit single-byte coded graphic character sets — Part 1: Latin alphabet No. 1* (referred to in ISO/IEC 29500 as the ANSI character set)

ISO/IEC 9075-1, *Information technology — Database languages — SQL — Part 1: Framework (SQL/Framework)*

ISO/IEC 10118-3:2004, *Information technology — Security techniques — Hash-functions — Part 3: Dedicated hash-functions*.

ISO/IEC 10646, *Information technology — Universal Coded Character Set (UCS)*.

ISO/IEC 14496-22:2009, *Information technology — Coding of audio-visual objects — Part 22: Open Font Format*

Japanese Industrial Standard, JIS X 0301: *Data elements and interchange formats —Information interchange — Representation of dates and times*. Japan, 2002.

Kingdom of Saudi Arabia, Ministry of Islamic Affairs, Endowments, Da'wah and Guidance.

Korean Law Enactment No. 4, 1961.

Faure, D. (n.d.). *Creating and Using Components (KParts)*. http://techbase.kde.org/Projects/Documentation.

Maimon, Rabbi Moshe ben, *Complete Restatement of the Oral Law (Mishneh Torah)*.

Ausbrooks, Ron, et al. *Mathematical Markup Language (MathML) Version 2.0 (Second Edition)*. October 21, 2003. http://www.w3.org/TR/MathML/.

Kaliski, B. *The MD2 Message-Digest Algorithm*. April 1992. http://www.ietf.org/rfc/rfc1319.txt

Rivest, R. *The MD4 Message-Digest Algorithm*. April 1992. http://www.ietf.org/rfc/rfc1320.txt

The MD5 Message-Digest Algorithm. April 1992. http://www.ietf.org/rfc/rfc1321.txt.

National Measurement Regulations 1999, Commonwealth of Australia
http://www.comlaw.gov.au/Details/F2011C00445

NIST Guide to SI Units, http://physics.nist.gov/Pubs/SP811/appenB9.html

QuickTime File Format Specification (2007-09-04 version)
http://developer.apple.com/standards/classicquicktime.html

Resource Description Framework (RDF), http://www.w3.org/RDF/

RFC 822, *Standard for ARPA Internet Text Messages* (http://www.ietf.org/rfc/rfc0822.txt)

RFC 2045, Borenstein, N., and N. Freed. *Multipurpose Internet Mail Extensions (MIME) Part One: Format of Internet Message Bodies*. The Internet Society. 1996. http://www.ietf.org/rfc/rfc2045.txt

RFC 2119, Bradner, Scott, 1997: *Key words for use in RFCs to Indicate Requirement Levels*. http://www.ietf.org/rfc/rfc2119.txt

RFC 2616, Berners-Lee, T., R. Fielding, H. Frystyk, J. Gettys, P. Leach, L. Masinter, and J. Mogul. *Hypertext Transfer Protocol—HTTP/1.1*. The Internet Society. 1999. http://www.ietf.org/rfc/rfc2616.txt

RFC 3066, Alvestrand, H. *Tags for the Identification of Languages*. The Internet Society. 2001. http://www.ietf.org/rfc/rfc3066.txt

RFC 3339, Klyne, G. and C. Newman. *Date and Time on the Internet: Timestamps*. The Internet Society. 2002. http://www.ietf.org/rfc/rfc3339.txt

RFC 3629, Yergeau, F. *UTF-8, a transformation format of ISO 10646*. The Internet Society. 2003.
http://www.ietf.org/rfc/rfc3629.txt

RFC 3986, Berners-Lee, T., R. Fielding, and L. Masinter. *Uniform Resource Identifier (URI): Generic Syntax*. The
Internet Society. 2005. http://www.ietf.org/rfc/rfc3986.txt

Simple Object Access Protocol (SOAP), http://www.w3.org/TR/soap12

SMIL, Bulterman, D., Grassel, G., Jansen, J., Koivisto, A., Layaïda, N., Michel, T., et al. (2005, December 13).
Synchronized Multimedia Integration Language (SMIL 2.1). Retrieved from W3C: http://www.w3.org/TR/SMIL/

SVG, Andersson, O., Armstrong, P., Axelsson, H., Berjon, R., Bézaire, B., Bowler, J., et al. (2003, January 14).
Scalable Vector Graphics (SVG) 1.1 Specification. Retrieved from W3C - World Wide Web Consortium:
http://www.w3.org/TR/SVG/

The GNOME Project. (2003, December 12). *Component Model - Bonobo Document Model*. Retrieved from The
GNOME Development Site: http://developer.gnome.org/bonobo-activation/stable/

The Unicode Consortium. *The Unicode Standard*, http://www.unicode.org/standard/standard.html.

Unicode Technical Note #28, *Nearly Plain-Text Encoding of Mathematics*. August 29, 2006,
http://www.unicode.org/notes/tn28

United States Postal Service. *Domestic Mail Manual*. United States Postal Service. November 8, 2007.
http://pe.usps.com/cpim/ftp/manuals/dmm300/Full/MailingStandards.pdf

The Units of Measurement Regulations 1995, United Kingdom
http://www.opsi.gov.uk/si/si1995/Uksi_19951804_en_2.htm

Universal Postal Union. *POST*CODE: Postal addressing systems*. Berne: UPU Publications, 2006, ISBN 92-95025-
37-7, ISSN 1020-6019

Web Accessibility Initiative (WAI), http://www.w3.org/WAI/

XSLT, Clark, James, *XSL Transformations (XSLT) Version 1.0*, World Wide Web Consortium Recommendation.
1999. http://www.w3.org/TR/xslt

XML, Tim Bray, Jean Paoli, Eve Maler, C. M. Sperberg-McQueen, and François Yergeau (editors). *Extensible
Markup Language (XML) 1.0, Fourth Edition*. World Wide Web Consortium. 2006.
http://www.w3.org/TR/2006/REC-xml-20060816/ [Implementers should be aware that a further correction of
the normative reference to XML to refer to the 5th Edition will be necessary when the related Reference
Specifications to which this International Standard also makes normative reference and which also depend upon
XML, such as XSLT, XML Namespaces and XML Base, are all aligned with the 5th Edition.]

XML Base, Marsh, Jonathan. *XML Base*. World Wide Web Consortium. 2001. http://www.w3.org/TR/2001/REC-
xmlbase-20010627/

XML Namespaces, Tim Bray, Dave Hollander, Andrew Layman, and Richard Tobin (editors). *Namespaces in XML 1.0 (Third Edition)*, 8 December 2009. World Wide Web Consortium. http://www.w3.org/TR/2009/REC-xml-names-20091208/

XPATH, Clark, James; DeRose, Steve *XML Path Language (XPath) Version 1.0*, World Wide Web Consortium Recommendation. 1999. http://www.w3.org/TR/xpath.

XML Schema Part 0: Primer (Second Edition), W3C Recommendation 28 October 2004, http://www.w3.org/TR/xmlschema-0/

XML Schema Part 1: Structures (Second Edition), W3C Recommendation 28 October 2004, http://www.w3.org/TR/xmlschema-1/

XML Schema Part 2: Datatypes (Second Edition), W3C Recommendation 28 October 2004, http://www.w3.org/TR/xmlschema-2/

.ZIP File Format Specification from PKWARE, Inc., version 6.2.0 (2004), as specified in http://www.pkware.com/documents/APPNOTE/APPNOTE_6.2.0.txt