INTERNATIONAL STANDARD

ISO/IEC 30114-2

First edition
2018-04

Information technology — Extensions of Office Open XML file formats —

Part 2:
Character repertoire checking

Technologies de l'information — Extensions de formats de fichiers Office Open XML —

Partie 2: Vérification de répertoire de caractères

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.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

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. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see the following URL: www.iso.org/iso/foreword.html.

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

A list of all parts in the ISO/IEC 30114 series can be found on the ISO website.

Introduction

ISO/IEC 29500 was designed to allow the addition of markup and other data to Open Office XML (OOXML) documents, and to allow OOXML applications unaware of such markup and date to provide reasonable results.

This document standardizes a set of extensions for character repertoire checking. For example, it can be used to ensure that some cells in OOXML Spreadsheet documents contain characters in ISO/IEC 8859-1[1] only.

Information technology — Extensions of Office Open XML file formats —

Part 2:
Character repertoire checking

1   Scope

This document provides an extension of OOXML for embedding CREPDL scripts within an OOXML document, and validating character contents in the OOXML document against these CREPDL scripts.

2   Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements 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 19757-2, *Information technology — Document Schema Definition Language (DSDL) — Part 2: Regular-grammar-based validation — RELAX NG*

ISO/IEC 19757-7, *Information technology — Document Schema Definition Languages (DSDL) — Part 7: Character Repertoire Description Language (CREPDL)*

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

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

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

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

W3C XML, Extensible Markup Language (XML) 1.0, W3C Recommendation, available at http://www.w3.org/TR/xml

W3C XML Schema Part 1: Structure, W3C Recommendation, available at http://www.w3.org/TR/xmlschema-1/

3   Terms and definitions

For the purposes of this document, the terms and definitions given in W3C XML, W3C XML Schema Part 1, ISO/IEC 19757-2, ISO/IEC 19757-7, ISO/IEC 29500-1, ISO/IEC 29500-2, ISO/IEC 29500-3, and ISO/IEC 29500-4 and the following apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— IEC Electropedia: available at http://www.electropedia.org/

— ISO Online browsing platform: available at http://www.iso.org/obp

3.1
CREPDL part
OPC part containing a CREPDL script

3.2
OPC part
part

Note 1 to entry: Parts are defined in ISO/IEC 29500-2.

Note 2 to entry: Since "part" is too general, this document uses "OPC part", instead.
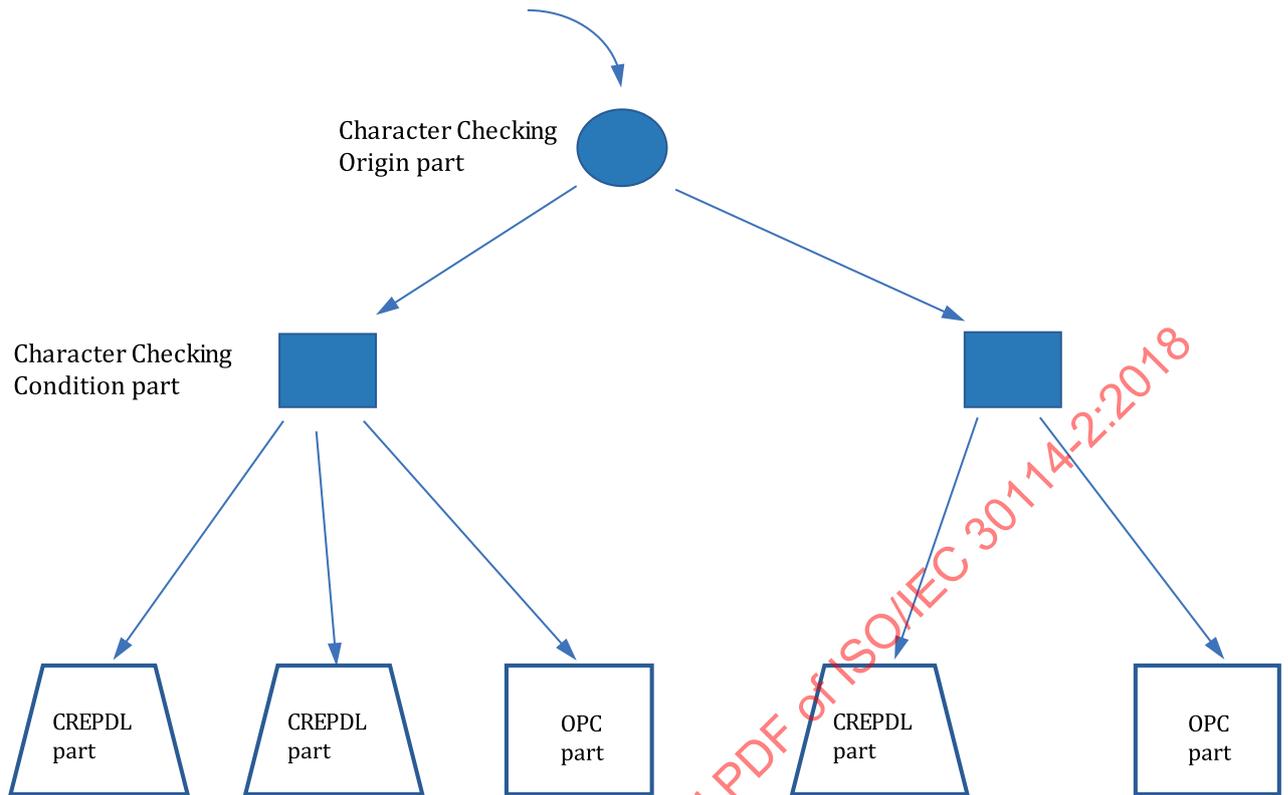
4   Syntax

4.1   General

This document introduces the Character Checking Origin part and Character Checking Condition parts. OOXML documents, which shall conform to ISO/IEC 29500-1 or ISO/IEC 29500-4, may contain these OPC parts for embedding CREPDL scripts. Embedded CREPDL scripts are also OPC parts. These OPC parts shall conform to ISO/IEC 29500-2, and CREPDL scripts shall conform to ISO/IEC 19757-7. Relationship names and content types relating to these OPC parts are defined in Annex B.

NOTE   ISO/IEC TR 30114-1[2] lists two extensibility mechanisms, ISO/IEC 29500-3 MCE (Markup Compatibility and Extensibility) and foreign OPC parts, for adding markup and other data to OOXML documents. This document uses foreign OPC parts rather than MCE.

When a CREPDL script contains a reference to a CREPDL script in another part, both the source and the referenced parts shall be the targets of relationships from the same Character Checking Condition part.

Figure 1 shows a package with the Character Checking Origin part and two Character Checking Condition parts. The example Character Checking Origin part references two Character Checking Condition parts, each of which references a CREPDL script as well as a part whose character contents are validated against the CREPDL script.

Figure 1 — Character Checking Origin part and Character Checking Condition parts

4.2 Character Checking Origin part

The Character Checking Origin part shall be an empty stream of the well-defined media type specified in Annex B.

An OOXML document shall have at most one Character Checking Origin part. If it exists, it shall be targeted from the package using the well-defined relationship type specified in Annex B.

4.3 Character Checking Condition part

4.3.1 General

A Character Checking Condition part shall be an XML document, which shall conform to W3C XML. It shall be of the well-defined media type specified in Annex B. A Character Checking Condition part may contain markup in the Markup Compatibility namespace as specified in ISO/IEC 29500-3. After the removal of any extensions by an MCE processor, which shall conform to ISO/IEC 29500-3, a Character Checking Condition part shall be valid against the schemas in Annex A and shall further satisfy requirements in the rest of 4.3. Validity against the schema in A.1 shall conform to W3C XML Schema Part 1, and that against the schema in A.2 shall conform to ISO/IEC 19757-2.

A document of Office Open XML conformance class Strict or Transitional may have one or more Character Checking Condition parts. Each Character Checking Condition part shall be targeted from the Character Checking Origin part using the well-defined relationship type specified in Annex B.

EXAMPLE

```
<Condition xmlns="http://schemas.openxmlformats.org/
  officeDocumentExtension/2015/characterCheckingConstraint"
xmlns:r="http://schemas.openxmlformats.org/officeDocument/2006/relationships">
  <RootScript r:id="rId1"/>
  <Location r:id="rId2"/>
</Condition>
```

4.3.2 Condition element

A `Condition` element is the root element of a Character Checking Condition part.

A `Condition` element is of the namespace `http://schemas.openxmlformats .org/officeDocumentExtension/2015/characterCheckingConstraint`. A `Condition` element has a `RootScript` element and zero or more Location elements in this sequence. It has no attributes.

4.3.3 RootScript element

A `RootScript` element is of the namespace `http://schemas.openxmlformats.org/ .officeDocumentExtension/2015/characterCheckingConstraint`. A `RootScript` element has the `r:id` attribute and no child elements.

The `r:id` attribute specifies the id of the relationship to the root CREPDL script. This relationship shall be one of the relationships from the Character Checking Condition part to CREPDL scripts.

4.3.4 Location element

A `Location` element specifies a location in the referenced part. Character contents at the specified location are subject to character checking.

A `Location` element is of the namespace `http://schemas.openxmlformats .org/officeDocumentExtension/2015/characterCheckingConstraint`. A Location element has the `r:id` attribute and no child elements. The `r:id` attribute specifies the id of the relationship to the part whose character contents are validated.

4.4 CREPDL part

A CREPDL part shall be an XML document of the media type specified in Annex B. A CREPDL part may contain markup in the Markup Compatibility namespace as specified in ISO/IEC 29500-3. After the removal of any extensions by an MCE processor as specified in ISO/IEC 29500-3, a CREPDL part is a CREPDL script as specified in ISO/IEC 19757-7.

A document of conformance class Office Open XML Strict or Transitional may have one or more CREPDL parts. Some CREPDL parts are directly referenced by `r:id` attributes of `RootScript` elements of Character Checking Condition parts. Other CREPDL parts are referenced directly or indirectly by these CREPDL parts by the href attribute of ref elements. Each CREPDL part shall be the target of a relationship from one or more Character Checking Condition parts using the relationship type specified in Annex B.

5   Processing model

a)   Visit each Character Checking Condition part and apply the MCE processor to it.

b)   Retrieve the CREPDL script referenced by the `RootScript` element in this Character Checking Condition part, and apply the MCE processor to it.

c)   Against this CREPDL script, for each `Location` element in this Character Checking Condition part, validate the character contents of the part referenced by this `Location` element.

Annex A
(normative)

Schemas for Character Checking Constraint parts

A.1 W3C XML Schema

```
<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema"
elementFormDefault="qualified"
targetNamespace="http://schemas.openxmlformats.org/officeDocumentExtension/
2015/characterCheckingConstraint"
xmlns:r="http://purl.oclc.org/ooxml/officeDocument/relationships"
xmlns:c="http://schemas.openxmlformats.org/officeDocumentExtension/
2015/characterCheckingConstraint">
  <xs:import namespace="http://purl.oclc.org/ooxml/officeDocument/relationships"
schemaLocation="shared-relationshipReference.xsd"/>
  <xs:element name="Condition">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="c:RootScript"/>
        <xs:element minOccurs="0" maxOccurs="unbounded" ref="c:Location"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="RootScript">
    <xs:complexType>
      <xs:attributeGroup ref="r:r_id"/>
    </xs:complexType>
  </xs:element>
  <xs:element name="Location">
    <xs:complexType>
      <xs:attributeGroup ref="r:r_id"/>
    </xs:complexType>
  </xs:element>
</xs:schema>
```

A.2 RELAX NG

```
default namespace =
  "http://schemas.openxmlformats.org/officeDocumentExtension/
2015/characterCheckingConstraint"
include "shared-relationshipReference.rnc"
start = element Condition { rootScript, location* }
rootScript = element RootScript { r_id}
location = element Location { r_id }
```