Date:   2012-07-02

ISO/IEC CD 30114-2

Secretariat:   JISC

Information technoology — Extensions of Office Open XML File Formats — Part 2: Character Repertoire Checking

*Technologies de l'information — Extensions des formats Office Open XML File — Partie 2: Répertoire de caractères Vérification*

Warning

This document is not an ISO International Standard. It is distributed for review and comment. It is subject to change without notice and may not be referred to as an International Standard.

Recipients of this draft are invited to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.

**Copyright notice**

This ISO document is a working draft or committee draft and is copyright-protected by ISO. While the reproduction of working drafts or committee drafts in any form for use by participants in the ISO standards development process is permitted without prior permission from ISO, neither this document nor any extract from it may be reproduced, stored or transmitted in any form for any other purpose without prior written permission from ISO.

Requests for permission to reproduce this document for the purpose of selling it should be addressed as shown below or to ISO's member body in the country of the requester:

[Indicate the full address, telephone number, fax number, telex number, and electronic mail address, as appropriate, of the Copyright Manger of the ISO member body responsible for the secretariat of the TC or SC within the framework of which the working document has been prepared.]

Reproduction for sales purposes may be subject to royalty payments or a licensing agreement.

Violators may be prosecuted.

Contents Page

Foreword iv

Introduction v

1 Scope 1

2 Conformance 1

2.1 Data Conformance 1

2.2 Application Conformance 1

3 Normative references 2

4 Terms and definitions 2

5 Syntax 2

5.1 General 2

5.2 Character Checking Origin part 3

5.3 Character Checking Condition part 3

5.3.1 Location element 3

5.4 CREPDL Part 3

6 Processing model 4

Annex A (normative) Schemas for Character Checking Constraint parts 5

A.1 W3C XML Schema 5

A.2 RELAX NG 5

Annex B (normative) Standard Namespaces, Content Types, and Relationship Types 6

B.1 Namespaces 6

B.2 Content Types 6

B.3 Relationship Types 6

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 30114‑2 was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information Technology*, Subcommittee SC 34, *Document Description and Processing Languages*.

ISO/IEC 30114 consists of the following parts, under the general title *Information technoology — Extensions of Office Open XML File Formats*:

* *Part 1: Guidelines*
* *Part 2: Character Repertoire Checking*

Introduction

ISO/IEC 29500 is designed to allow safe extensions. An OOXML extension is safe if OOXML documents containing this extension are guaranteed to conform to ISO/IEC 29500 and are thus guaranteed to be processable by conformant applications of ISO/IEC 29500. Safe extensions of ISO/IEC 29500 may or may not be standardized. ISO/IEC 30114 provides a collection of standardized safe extensions of OOXML.

This part of ISO/IEC 30114 standardizes a set of safe extensions for character repertoire checking. For example, it can be used to ensure that some cells in OOXML Spreadsheet documents contain characters in ISO 8859-1 only.

The structure of this part of ISO/IEC 301147 is as follows. ...

Information technoology — Extensions of Office Open XML File Formats — Part 2: Character Repertoire Checking

# Scope

This part of the multi-part standard provides an extension for embedding CREPDL scripts (as specified in 19757-7) as OPC parts within OOXML documents, associating such OPC parts with certain locations in OOXML documents, and examining whether or not text chunks at these locations are in the specified repertoires.

Note This draft covers two locations: sdtPr of WML and c of SML. Should we try to provide more? Here are some other possibilities:

* Comments and sheet names in SML
* The contents of an entire WML document
* The content of a bookmark (@w:id or @w:nam of a w:bookmarkStart w:bookmarkEnd pair)
* The content of a paragraph (the value of w:style/w:name/@w:val or w:style/@w:styleId)

# Conformance

## Data Conformance

An OOXML document shall be conformant to this part of the standard if it satisfies requirements stated in this part.

## Application Conformance

To be supplied.

# 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 19757‑2, *Document Schema Definition Languages (DSDL) — Regular-grammar-based validation — RELAX NG*

ISO/IEC 19757‑4, *Document Schema Definition Languages (DSDL) — Namespace-based validation dispatching language — NVDL*

ISO/IEC 19757‑7, *Document Schema Definition Languages (DSDL) — Character repertoire description language — CREPDL*

ISO/IEC 29500‑1, *Office Open XML File Formats — Fundamentals and Markup Language Reference*

ISO/IEC 29500‑2, *Office Open XML File Formats — Open Packaging Conventions*

ISO/IEC 29500‑3, *Office Open XML File Formats — Markup Compatibility and Extensibility*

ISO/IEC 29500‑4, *Office Open XML File Formats — Transitional Migration Features*

# Terms and definitions

For the purposes of this document, the terms and definitions given in the normative references and the following apply.

4.1

OPC part

Part as defined in 29500-2

# Syntax

## General

The character checking parts consist of the Character Checking Origin part and Character Checking Condition parts. Relationship names and content types relating to these parts are defined in Annex B.

Figure X-X 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 scirpt.

Editor's Note: This figure is very similar to Figure 12-1 in Part 2.

## Character Checking Origin part

The Character Checking Origin part shall be an empty stream of the content type application/vnd.openxmlformats-extension.character-checking-origin.

## Character Checking Condition part

A Character Checking Condition part shall be an XML document of the content type application/vnd.openxmlformats-extension.character-checking-origin. This XML document has a Condition element as the root. It has zero or more Location elements in this sequence.

Example:

<Condition

xmlns="http://schemas.openxmlformats.org/officeDocumentExtension/2015/characterCheckingConstraint">

<Location/>

</Condition>

### Location element

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

A location element has no elements and attributes, and is assumed to specify the entire part.

Note:

## 

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

from a Character Checking Condition part.

# Processing model

Visit each Character Checking Condition part in sequence. Check the character content at the specified location against the CREPDL script referenced by this Character Checking Condition part.

### 

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |

1. (normative)  
     
   Schemas for Character Checking Constraint parts
   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:c=

"http://schemas.openxmlformats.org/officeDocumentExtension/2015/characterCheckingConstraint">

<xs:element name="Condition">

<xs:complexType>

<xs:sequence>

<xs:element minOccurs="0" ref="c:Location"/>

</xs:sequence>

</xs:complexType>

</xs:element>

<xs:element name="Location">

<xs:complexType/>

</xs:element>

</xs:schema>

* 1. RELAX NG

default namespace =

"http://schemas.openxmlformats.org/officeDocumentExtension/2015/characterCheckingConstraint"

start = element Condition { location? }

location = element Location { empty }<?xml version="1.0" encoding="UTF-8"?>

1. (normative)  
     
   Standard Namespaces, Content Types, and Relationship Types
   1. Namespaces

Character Checking Condition part

http://schemas.openxmlformats.org/officeDocumentExtension/2015/characterCheckingConstraint

CREPDL script

"http://purl.oclc.org/dsdl/crepdl/ns/structure/1.0"

* 1. Content Types

Character Checking Condition part application/vnd.openxmlformats-extension.character-checking-condition

Character Checking Origin part application/vnd.openxmlformats-extension.character-checking-origin

CREPDL script application/xml

* 1. Relationship Types

From the package to the origin part

http://schemas.openxmlformats.org/officeDocumentExtension/relationships/2015/characterCheckingOrigin

From the origin part to a condition part

<http://schemas.openxmlformats.org/officeDocumentExtension/relationships/2015/characterCheckingConstraint>

From a condition part to a CREPDL script part

http://schemas.openxmlformats.org/officeDocumentExtension/relationships/2015/characterCheckingScript