INTERNATIONAL ISO/IEC

Third edition
2020-06

# Information technology — Document Schema Definition Languages (DSDL) —

## Part 3: Rule-based validation using Schematron

*Technologies de l'information — Langages de définition de schéma de documents (DSDL) —*

*Partie 3: Validation basée sur des règles à l'aide de Schematron*

Reference number
ISO/IEC 19757-3:2020(E)



ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

# 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.

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) or the IEC list of patent declarations received (see http://patents.iec.ch).

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

For an explanation of 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 www.iso.org/iso/foreword.html.

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

This third edition cancels and replaces the second edition (ISO/IEC 19757-3:2016), which has been technically revised.

The main changes compared to the previous edition are as follows:

— query language bindings have been added for XSLT 3.0 (Annex J) and XPath 3.0 (Annex K);

— annexes pertaining to XPath and XSLT query language bindings (Annexes H to K) are now all normative, while those for EXSLT (Annex L) and STX (Annex M) remain informative.

A list of all parts in the ISO/IEC 19757 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

# Introduction

ISO/IEC 19757 (all parts) defines a set of Document Schema Definition Languages (DSDL) that can be used to specify one or more validation processes performed against Extensible Markup Language (XML) or Standard Generalized Markup Language (SGML) documents. [XML is an application profile SGML (see ISO 8879).]

A document model is an expression of the constraints to be placed on the structure and content of documents to be validated with the model. A number of technologies have been developed through various formal and informal consortia since the development of Document Type Definitions (DTDs) as part of ISO 8879, notably by the World Wide Web Consortium (W3C) and the Organization for the Advancement of Structured Information Standards (OASIS). A number of validation technologies are standardized in DSDL to complement those already available as standards or from the industry.

Through the validation that a structured document conforms to specified constraints in structure and content, the potentially many applications acting on the document are relieved from duplicating the task of confirming that such requirements have been met. Historically, such tasks and expressions have been developed and utilized in isolation, without consideration of how the features and functionality available in other technologies can enhance validation objectives.

The main objective of ISO/IEC 19757 (all parts) is to bring together different validation-related tasks and expressions to form a single extensible framework that allows technologies to work in series or in parallel to produce a single or a set of validation results. The extensibility of DSDL accommodates validation technologies not yet designed or specified.

In the past, different design and use criteria have led users to choose different validation technologies for different portions of their information. Bringing together information within a single XML document sometimes prevents existing document models from being used to validate sections of data. By providing an integrated suite of constraint description languages that can be applied to different subsets of a single XML document, ISO/IEC 19757 (all parts) allows different validation technologies to be integrated under a well-defined validation policy.

The structure of this document is as follows. Clause 5 describes the syntax of an ISO Schematron schema. Clause 6 describes the semantics of a correct ISO Schematron schema; the semantics specify when a document is valid with respect to an ISO Schematron schema. Clause 7 describes conformance requirements for implementations of ISO Schematron validators. Annex A provides the ISO/IEC 19757-2 (RELAX NG) schema for ISO Schematron. Annex B provides the ISO Schematron schema for constraints in ISO Schematron that cannot be expressed by the schema of Annex A. Annex C provides the default query language binding to XSLT1. Annex D provides an ISO/IEC 19757-2 (RELAX NG compact syntax) schema and corresponding ISO Schematron schema for a simple XML language Schematron Validation Report Language. Annex E provides motivating design requirements for ISO Schematron. Annex F specifies certain Schematron elements to be used in external vocabularies. Annex G provides a simple example of a multi-lingual schema. Annexes H to M provide query language bindings. Annex N shows example usage of Schematron properties.

This edition is backwards compatible with ISO/IEC 19757-3:2016, supersedes it and provides extra query language bindings, in particular for XSLT3.

Considered as a document type, a Schematron schema contains natural-language assertions concerning a set of documents, marked up with various elements and attributes for testing these natural-language assertions and for simplifying and grouping assertions.

Considered theoretically, a Schematron schema reduces to a non-chaining rule system whose terms are Boolean functions invoking an external query language on the instance and other visible XML documents, with syntactic features to reduce specification size and to allow efficient implementation.

Considered analytically, Schematron has two characteristic high-level abstractions: the pattern and the phase. These allow the representation of non-regular, non-sequential constraints that ISO/IEC 19757-2 cannot specify and various dynamic or contingent constraints.

 

This document is based on the Schematron[2] assertion language. The `let` element is based on XCSL[4]. Other features arise from the half-dozen early open-source implementations of Schematron in diverse programming languages and from discussions in electronic forums by Schematron users and implementers.