Validation of Define-XML with XML Schema and Schematron

Lex Jansen
2025-11-03

Validation of Define-XML

Validation of Define-XML has 3 phases:

1. **Does the document follow the XML ground rules (*well-formed*)?**
   - A single element contains all other elements
   - Properly nested tags
   - Properly quotes attributes
   - No illegal characters (correct encoding)
   - This is an absolute pre-condition for the next 2 phases.

2. **Does the document follow the specific Define-XML grammar?**
   - Names of elements and attributes that must/may be included
   - Order of elements
   - Cardinality of elements
   - Datatypes of attributes
   - Enumerations (valid values)
   - Uniqueness constraints

3. **Does the data make sense (*business rules*)**

Define-XML Specification, mostly implemented by XML Schema

Define-XML Specification

Validating XML - Define-XML Conformance Rules

Rule Identifier	Displayed Link	Plain Text Rule	Rule	Rule Message	Applicable Versions	Source Type	XPaths	Element	Attribute
1	\\ALL elements 	Elements must be ordered in accordance with the global element ordering rules in the Define-XML specification.	//\<element OID> is not allowed under [parent element]/@OID.	Element [element]/@OID s not allowed under [parent element]/@OID.	Both 2.0 and 2.1	Schema		ALL elements	
2	\\ODM 	Element ODM must be provided in any Define-XML document.	/ODM must be provided.	Element ODM is missing in the Define-XML document.	Both 2.0 and 2.1	Specification	/ODM	ODM	
3	\\ODM 	No more than one ODM element may be provided in the Define-XML document.	No more than one ODM element may be provided.	Element ODM is provided more than once in the Define-XML document.	Both 2.0 and 2.1	Specification	/ODM	ODM	
4	\<a href="\<a href="https://wiki.cdisc.org/x/Z6OzAg"target="_blank">ODM	The value of attribute CreationDateTime must be in ISO 8601 datetime format	The value of @CreationDateTime must be in ISO 8601	The value of CreationDateTime[@CreationDateTime] is not in ISO	Both 2.0 and 2.1	Schema	/ODM	ODM	CreationDateTime

Validating XML - Define-XML Conformance Rules

Usage	

Usage	• Requirement: This is populated with one of 3 values:
	   ○ "Required" when at least 1 instance of the element is required
	   ○ "Optional" when the element is optional
	   ○ "Conditional" when at least 1 instance of the element is required under certain conditions. It will include the conditions under which the element is Required.
	• Cardinality: This indicates the number of instances expected (e.g. "Exactly one", "One or more")
	• Business Rule(s): Beyond what is required for schema validation, these represent conformance rules that must be satisfied for a Define-XML document to be considered compliant with this specification.
	• Other Information: This is populated with any other information about the element, including the conditions under which the element is included, how the schema is applied to support the model, relative position of the element in the model, and so on.

Validating XML - Define-XML Business Rules
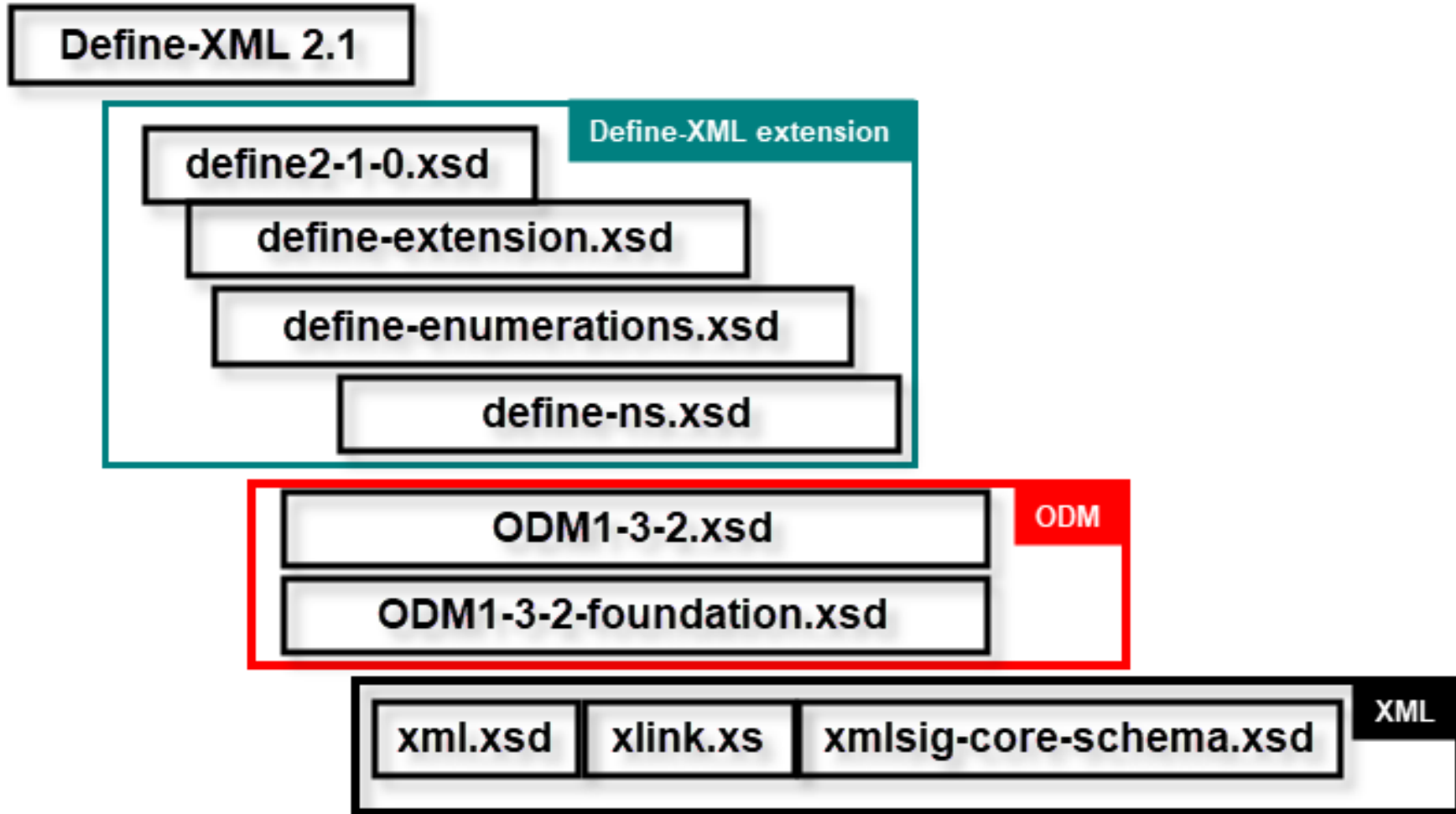
- Business Rule: StandardOID must match the OID for a def:Standard element with Type="IG".

- Business Rule: If this value is provided for any EnumeratedItem, it must be provided for all. **Note** that the Rank attribute does not imply a display order.

- Business Rule: For analysis datasets, if the ItemGroupDef def:IsNonStandard attribute is used, the def:Class should not be provided.

Codelist Name	CDISC Submission Value	
ADaM Basic Data Structure Subclass	BDSSC	
ADaM Basic Data Structure Subclass	NON-COMPARTMENTAL ANALYSIS	
ADaM Basic Data Structure Subclass	POPULATION PHARMACOKINETIC ANALYSIS	
ADaM Basic Data Structure Subclass	TIME-TO-EVENT	

Validating XML

- XML schema languages:
  - Document Type Definitions (DTD)
  - XML Schema (W3C)
  - Schematron
  - XQuery
  - RELAX NG

XML Schema for Define-XML v2.1

Define-XML 2.1

Define-XML extension
- define2-1-0.xsd
- define-extension.xsd
- define-enumerations.xsd
- define-ns.xsd

ODM
- ODM1-3-2.xsd
- ODM1-3-2-foundation.xsd

XML
- xml.xsd
- xlink.xs
- xmlsig-core-schema.xsd

Limitations

- XML Schema 1.0 provides no facilities to state that the value or presence of **one data item** is **dependent** on the values or presence of **other data items** (so-called *co-occurrence constraints*).
- Error messages are not very human friendly

Schematron

- Can compare arbitrary nodes in an XML file
- Namespace awareness
- Customizable friendly error messages
- Verify co-occurrence constraints
  - In a submission, if the ItemDef/def:Origin/@Type attribute is "Collected" then there has to be a def:DocumentRef child element
  - If the Purpose attribute is "Tabulations" then the def:Class/@Name attribute has to be "FINDINGS", "EVENTS", "INTERVENTIONS", ...
- Based on XSL transformations
- Everything not forbidden is allowed
- Schematron makes assertions about the document content

Validating XML - Schematron

```xml
<ODM
  xmlns="http://www.cdisc.org/ns/odm/v1.3"
  xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:def="http://www.cdisc.org/ns/def/v2.1"
  ODMVersion="1.3.2"
  FileOID="www.cdisc.org/StudyMSGv2/1/Define-XML_2.1.0"
  FileType="Snapshot"
  CreationDateTime="2021-01-24T16:59:33"
  def:Context="Submission"
  >
```

```xml
<ItemDef OID="IT.AE.AEREL" Name="AEREL" DataType="text" Length="16" SASFieldName="AEREL">
  <Description>
    <TranslatedText xml:lang="en">Causality</TranslatedText>
  </Description>
  <CodeListRef CodeListOID="CL.AEREL"/>
  <def:Origin Type="Collected" Source="Investigator" />
</ItemDef>
```

```xml
<sch:pattern id="ItemDef_Origin_Collected">
  <sch:title>Collected item needs CRF</sch:title>
  <sch:rule context="//odm:ItemDef/def:Origin[@Type='Collected']">
    <sch:let name="ODMCONTEXT" value="/odm:ODM/@def:Context"/>
    <sch:let name="NAME" value="../@Name"/>
    <sch:let name="OID" value="../@OID"/>
    <sch:let name="SOURCE" value="@Source"/>
    <sch:report test="(not(def:DocumentRef/def:PDFPageRef) and ($SOURCE='Investigator' or $SOURCE='Subject')) or not($ODMCONTEXT = 'Submission')"
      flag="ERROR" see="[Section 5.3.12.3 def:Origin Element](https://wiki.cdisc.org/x/f6OzAg) def:Origin">ItemDef element
      <sch:value-of select="$NAME"/> (OID=&quot;<sch:value-of select="$OID"/>&quot;)
      is Collected, with a source of 'Investigator' or 'Subject', but does not have a reference to a PDF document.</sch:report>
  </sch:rule>
</sch:pattern>
```
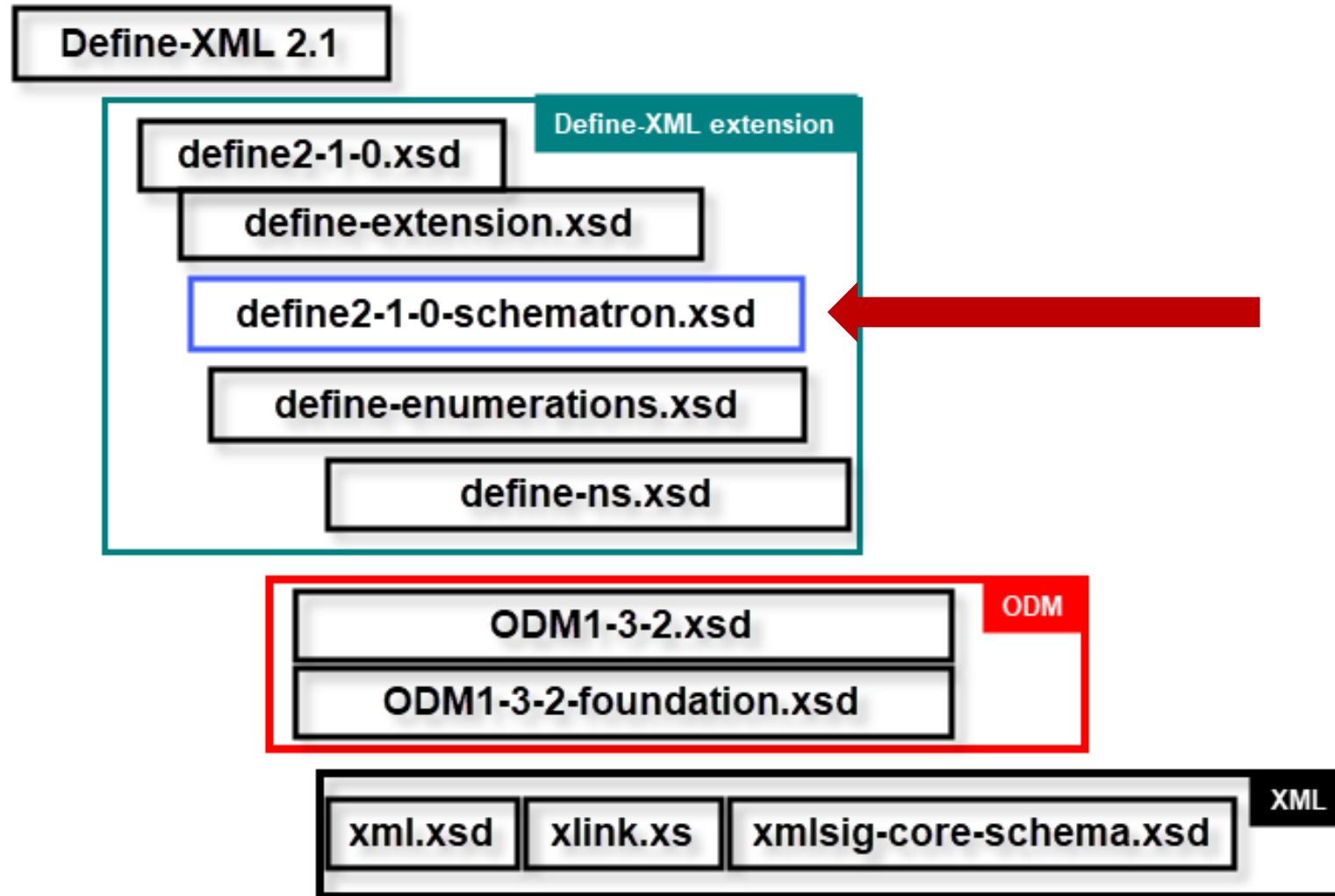
Rule ID	Severity	Reference	Message
ItemDef_Origin_Collected	ERROR	Section 5.3.12.3 def:Origin Element def:Origin	ItemDef element AEACN (OID="IT.AE.AEACN") is Collected, with a source of 'Investigator' or 'Subject', but does not have a reference to a PDF document.
ItemDef_Origin_Collected	ERROR	Section 5.3.12.3 def:Origin Element def:Origin	ItemDef element AEREL (OID="IT.AE.AEREL") is Collected, with a source of 'Investigator' or 'Subject', but does not have a reference to a PDF document.

"Assertions" and "Reports"

XML Schema for Define-XML v2.1 + Schematron

XML Schema for Define-XML v2.1 + Schematron

ItemGroupDef_Class_SubClass_Analysis	Only Analysis datasets have subclasses.	1
ItemGroupDef-Class-SubClass	Only certain SubClass Name values are allowed within a Class	1
ItemDef_Origin_Collected	Collected item needs CRF	2

Schematron Issue Details

Rule ID	Severity	Reference	Message
ODM_FileType	ERROR	Section 5.3.3 ODM Element ODM/@FileType	The FileType attribute on the ODM element is "Transactional ", but must have the value "Snapshot".
Standard_Not_Referenced	WARNING	Section 5.3.6 def:Standards Element	def:Standard element with OID="STD.2_1" (SDTMIG-MD, 1.1,) is not being referenced.
Standard_Not_Referenced	WARNING	Section 5.3.6 def:Standards Element	def:Standard element with OID="STD.4" (CDISC/NCI, 2020-12-18, DEFINE-XML) is not being referenced.
ItemGroupDef_NoData_ArchiveLocationID	ERROR		def:ArchiveLocationID on ItemGroupDef with OID 'IG.NV' should not be there when there is no data (def:HasNoData="Yes").
ItemGroupDef_ArchiveLocationID	ERROR		def:ArchiveLocationID on ItemGroupDef with OID 'IG.OE' is required when there is data.
ItemGroupDef_Standard_NonStandard_both	ERROR		ItemGroup 'IG.OE' should either reference a standard or be NonStandard (@def:IsNonStandard='Yes'), but not both.
ItemGroupDef_Standard_NonStandard	ERROR		ItemGroupDefs "QSPH" (IG.QSPH) should either reference a standard or be NonStandard (@def:IsNonStandard='Yes').
ItemGroupDef_Alias	ERROR		ItemGroupDefs "SUPPDM" (IG.SUPPDM) does not have an Alias with a DomainDescription, while ItemGroup represents a Supplemental Qualifiers dataset.
ItemGroupDef_Class_Tabulation	ERROR		ItemGroupDef "DM" (IG.DM) has invalid Class value "SUBJECT LEVEL ANALYSIS DATASET". For Tabulation datasets the allowed values are: EVENTS, FINDINGS, FINDINGS ABOUT, INTERVENTIONS, RELATIONSHIP, SPECIAL PURPOSE, STUDY REFERENCE, TRIAL DESIGN.
ItemGroupDef_Class_SubClass_Analysis	ERROR		ItemGroupDef "IG.AE", Class "AE" has SubClass value "ADVERSE EVENT", but is no Analysis dataset.
ItemGroupDef-Class-SubClass	ERROR		In ItemGroupDef "AE" (IG.AE), SubClass 'ADVERSE EVENT' can only occur within the Class 'OCCURRENCE DATA STRUCTURE'.
ItemDef_Origin_Collected	ERROR	Section 5.3.12.3 def:Origin Element def:Origin	ItemDef element AEACN (OID="IT.AE.AEACN") is Collected, with a source of 'Investigator' or 'Subject', but does not have a reference to a PDF document.
ItemDef_Origin_Collected	ERROR	Section 5.3.12.3 def:Origin Element def:Origin	ItemDef element AEREL (OID="IT.AE.AEREL") is Collected, with a source of 'Investigator' or 'Subject', but does not have a reference to a PDF document.

XML Schema Validation Report

Row	Column	XML Schema Rule	Description
12	67	cvc-enumeration-valid	Value 'Transactional ' is not facet-valid with respect to enumeration '[Snapshot, Transactional]'. It must be a value from the enumeration.
12	67	cvc-attribute.3	The value 'Transactional ' of attribute 'FileType' on element 'ODM' is not valid with respect to its type, 'FileType'.
3307	54	cvc-enumeration-valid	Value 'CRF' is not facet-valid with respect to enumeration '[Assigned, Collected, Derived, Not Available, Other, Predecessor, Protocol]'. It must be a value from the enumeration.
3307	54	cvc-attribute.3	The value 'CRF' of attribute 'Type' on element 'def:Origin' is not valid with respect to its type, 'OriginType'.

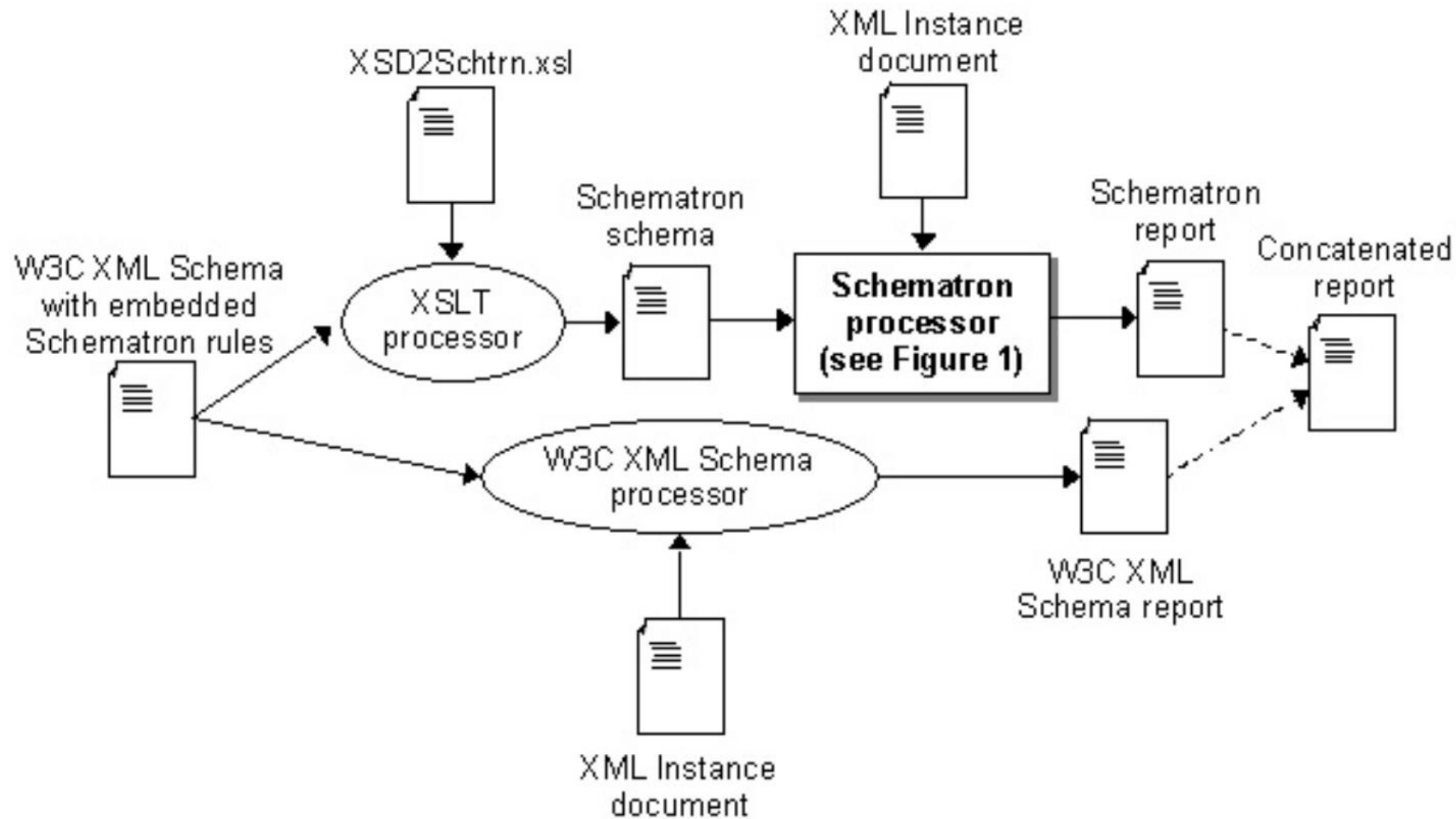# XML Schema for Define-XML v2.1 + Schematron - Technology

Figure 2: Processing a W3C XML Schema with embedded Schematron rules

- https://www.schematron.com/

  A website for all things Schematron

- https://codeberg.org/SchXslt/schxslt

  SchXslt [ʃˈɛksl̩t] – An XSLT-based Schematron processor

- https://github.com/lexjansen/schematron4define