

Documentation Quality Assurance with ISO Schematron

Octavian Nadolu, Syncro Soft
octavian_nadolu@oxygenxml.com
@OctavianNadolu

© 2019 Syncro Soft SRL. All rights reserved.



schematron
Structured
editing
XML
review
XQuery
Publish
PDF
DTD DocBook
oxygen
authoring
XML Editor
XSD SCHXSD
XPRRNC FO
frameworks
Profiling
WSDL
styles
visual
WebHelp
DITA
TEI
XSL
PHP
Ant
Js
JS
KML
XSLT
SVN
JSON
SVG
IDREES
WebDAV
Single
Sourc
Datab
XHTML
Cha
Col
We

About me

Software Architect at Syncro Soft

octavian.nadolu@oxygenxml.com

- 15+ years of XML technology experience
- Contributor for various XML-related open source projects
- Editor of Schematron QuickFix specification developed by a W3C community group

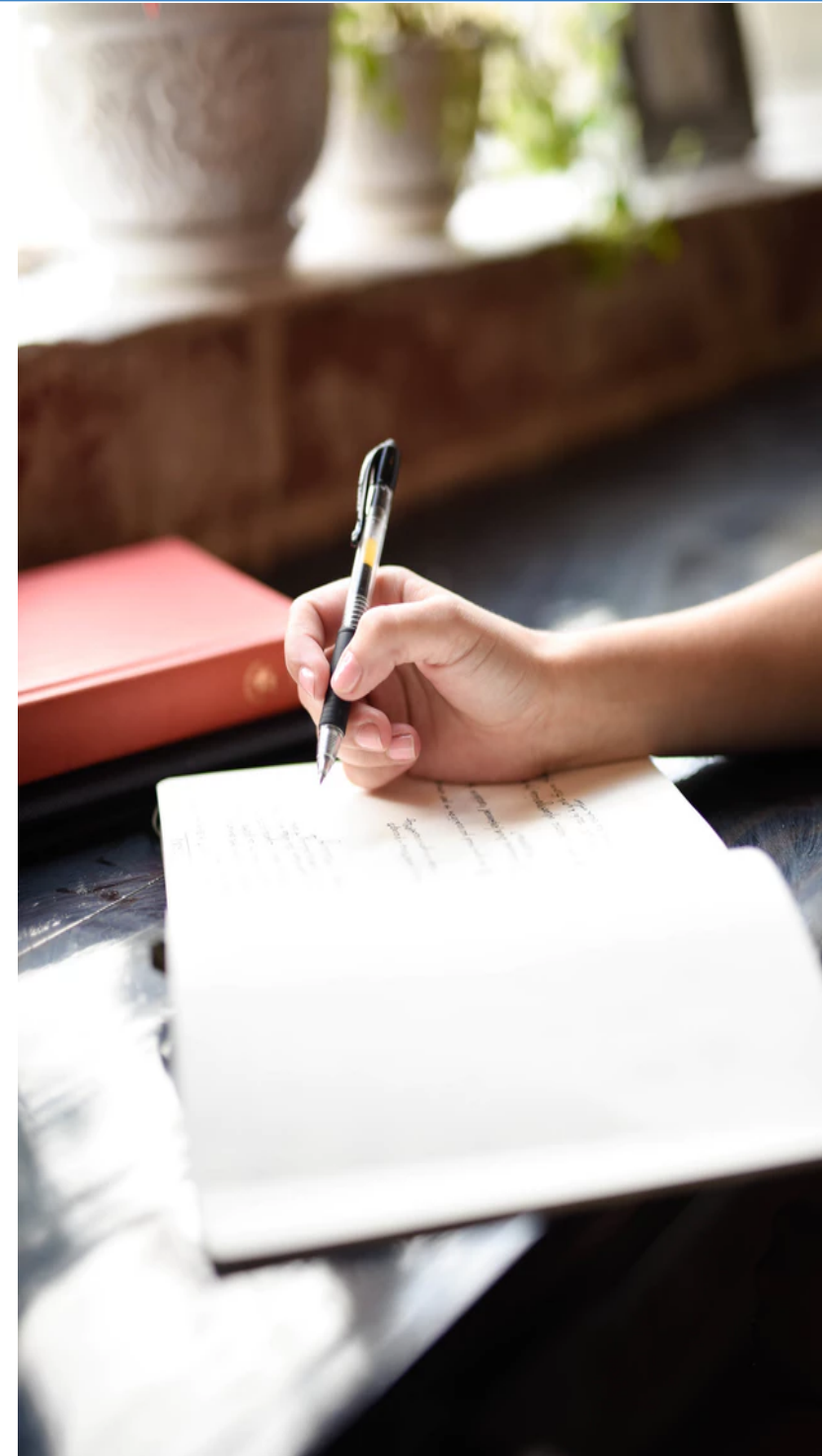


schematron
Structured
editing
XML
review
XQuery
Publish
PDF
IDREFS
WebDAV
DTD DocBook
oxygen
authoring
XML Editor
XSD SCHXSD Single
XPRRNC FO
frameworks
Profiling
WSDL
styles
visual
WebHelp
DITA
TEI
XSL
PHP
Ant
Js

JS
KML
XSLT
SVN
JSON
SVG
WebDAV
DocBook
authoring
XML Editor
Single
Source
Datab
XHTML
Cha
Col
We

Agenda

- Quality assurance (QA)
- ISO Schematron
- Schematron rules for a documentation project
- Types of quality assurance rules
- Generate a report



Quality Assurance

A way of **preventing mistakes** and **defects** in manufactured products



What is Schematron?

A natural language for making assertions in documents



Schematron is an ISO Standard

Schematron is an ISO standard adopted by hundreds of major projects in the past decade





Schematron History

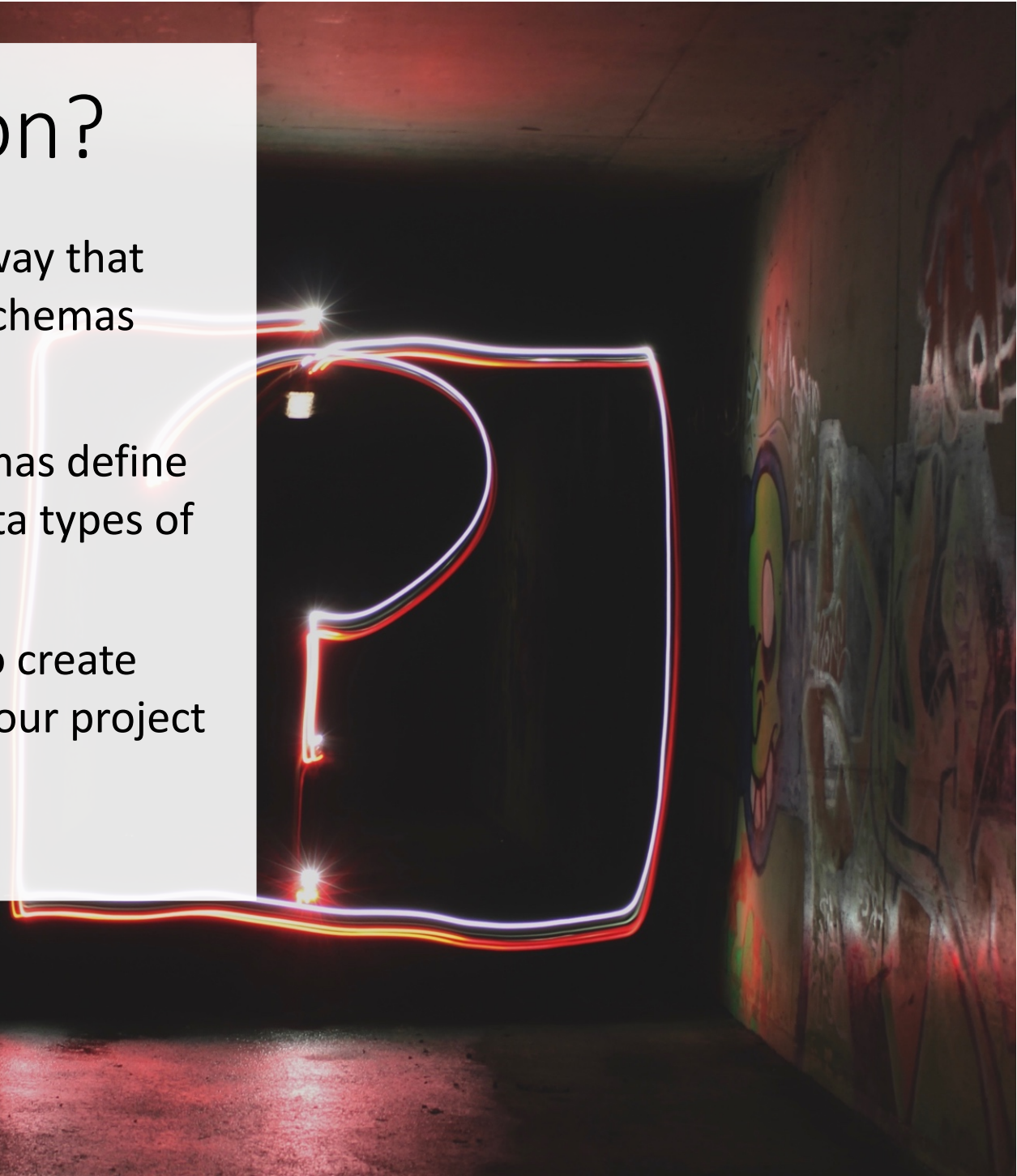
Schematron was invented by Rick Jelliffe

- Schematron 1.0 (1999), 1.3 (2000), 1.5 (2001)
- ISO Schematron (2006, 2010, 2016)

Why Schematron?

You can express constraints in a way that you cannot perform with other schemas (like XSD, RNG, or DTD).

- XSD, RNG, and DTD schemas define structural aspects and data types of the XML documents
- Schematron allows you to create custom rules specific to your project

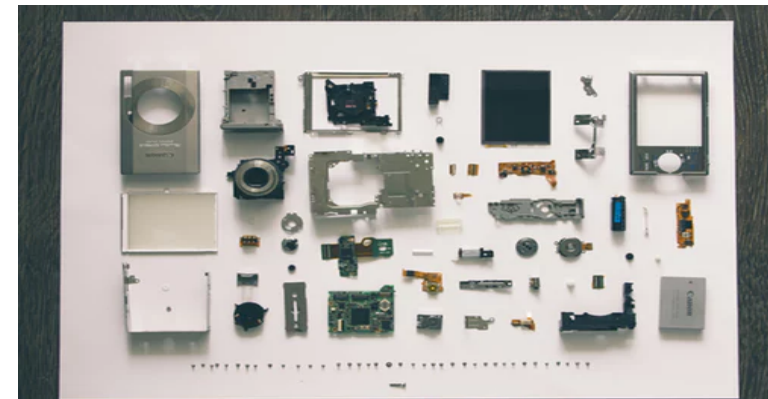


Schematron Usage

- Verify data inter-dependencies
- Check data cardinality
- Perform algorithmic checks

Used in Multiple Domains

- Financial
- Insurance
- Government
- Technical publishing



Schematron is Very Simple

There are only 6 basic elements:

assert

report

rule

pattern

schema

ns



Example of Rules

- **Titles** must be uppercase
- **Short descriptions** should not exceed 50 characters
- Avoid having **empty elements**
- Avoid having a semi-colon at the end of a **list item**
- **Lists** should contain more than one item



Oxygen User Guide

Quality assurance rules used in Oxygen user guide project



Introduction

Welcome to the User Manual of Oxygen XML Editor 21.1

Getting Started

Information and resources to help you get started using Oxygen XML Editor as quickly as possible

Installation

The requirements and installation instructions for each platform

Configuration

Description of all of the options that allow you to configure Oxygen XML Editor

Schematron Validation Result



Type of Problems

- Different type of problems
 - Code smell
 - Bug
 - Vulnerability
- Different severity
 - Fatal
 - Error
 - Warn
 - Info



Schematron Messages Severity

Severity level can be set in the value of the `@role` attribute from an `assert` or `report` element (possible values: `fatal`, `error`, `warn`, `info`)

```
<sch:rule context="related-links/linklist">  
  <sch:assert test="title" role="warn">  
    The linklist should have a title</sch:assert>  
</sch:rule>
```

Validation Result in Editing Framework

The screenshot displays an editing framework interface. The top panel shows a document structure with two unordered lists. The first list, located at `rulesAdvanced.sch#144`, contains four items: "Item 1", "Item 2", "Item 3", and "Item 4". The second list, located at `rulesAdvanced.sch#164`, contains three items: "1. Item 1", "2. Item 2", and "1. Item 3". A warning icon (lightbulb) is visible next to the first list.

Below the document view, a message reads: "Two consecutive unordered lists. You can probably merge them into one." The interface includes tabs for "Text", "Grid", and "Author".

The bottom panel shows a "Problems" window with a table of validation errors:

Info	Description - 20 items
!	W [ISO Schematron] Dynamically scaled images are not properly displayed, you should scale the image with an image
!	W [ISO Schematron] List items should not end with a semi-colon (;). If it is a sentence then end it with a full stop (.), c
!	W [ISO Schematron] List items should not end with a semi-colon (;). If it is a sentence then end it with a full stop (.), c
!	E [ISO Schematron] The indexterm element should be in a prolog.
!	E [ISO Schematron] Link text is same as @href attribute value. Please remove.
!	E [ISO Schematron] Image without a reference.
!	W [ISO Schematron] Two consecutive unordered lists. You can probably merge them into one.

The "Problems" window has a close button (X) and a refresh icon (gear).

Validation Result as HTML

Export validation result in an HTML file

Severity	Description	File	Location
error	The indexterm element should be in a prolog.	C:\samples.dita	Start line 71:22
error	Link text is same as @href attribute value. Please remove.	C:\samples.dita	Start line 32:84
error	Image without a reference.	C:\samples.dita	Start line 37:47
warning	Two consecutive unordered lists. You can probably merge them into one.	C:\samples.dita	Start line 43:33
warning	Two consecutive ordered lists. You can probably merge them into one.	C:\samples.dita	Start line 57:37
warning	Lines (2 - 124,) in codeblocks should not exceed 90 characters.	C:\samples.dita	Start line 135:50
warning	Possible XML Codeblock without @outputclass set to it.	C:\samples.dita	Start line 78:22
warning	Try to avoid inserting two consecutive notes with the same type.	C:\samples.dita	Start line 86:32
error	Empty element. Most DITA elements should not be empty.	C:\samples.dita	Start line 12:11
warning	All sections should have an @id attribute	C:\samples.dita	Start line 93:16
warning	The title is too long (78 chars). It should be less than 75 characters.	C:\samples.dita	Start line 100:18
warning	"menucascade" should contain more than one "uicontrol"	C:\samples.dita	Start line 106:33
warning	The figure should have a title.	C:\samples.dita	Start line 113:17
warning	The section should have a title.	C:\samples.dita	Start line 119:35
error	The text in a section element should be in a paragraph.	C:\samples.dita	Start line 119:35
error	Should use product key instead!	C:\samples.dita	Start line 130:12

Validation Result as SVRL

- As an XML file describing the validation errors, normally in Schematron Validation Reporting Language (SVRL)

```
<svrl:failed-assert test="matches(@href, '^http(s?)://')" role="warn"
    location="/topic[1]/body[1]/section[1]/p[3]/xref[1]">
  <svrl:text>An external link should start with http(s).</svrl:text>
</svrl:failed-assert>
```

Generate HTML from SVRL

SVRL report

Errors: 5

- **Text: Referenced resource "<http://en.wikipedia.org/wiki/Compost>" needs to have the "format" attribute set to it.**
 - Test: @format
 - Location: /topic[1]/body[1]/section[1]/ul[1]/li[3]/xref[1]
- **Text: An external link should start with http(s).**
 - Test: matches(@href, '^http(s?)://')
 - Location: /topic[1]/body[1]/section[1]/p[3]/xref[1]
- **Text: Cells are missing. (The number of cells for each row must be 3)**
 - Test: \$minColumnsNo >= \$reqColumnsNo
 - Location: /topic[1]/body[1]/section[2]/table[1]
- **Text: Ordered lists are not allowed, use unordered lists instead.**
 - Test: false()
 - Location: /topic[1]/body[1]/section[3]/ol[1]
- **Text: A list must have more than one item**
 - Test: count(li) > 1
 - Location: /topic[1]/body[1]/section[3]/ol[1]/li[4]/ul[1]

HTML Report – DE GRUYTER

I1

J1 ▾

The value '06' does not consist of one or two digits in element <month>. The value shouldn't start with a leading zero. (See Guidelines, section 'Publication Date')

Conversion step: validate-business-rules

02 06 2015 J2 ▾ 24 08 2015

permissions

S1 ▾

copyright-statement

© 2015 Boleslaw T. Karwowski, published by De Gruyter Open.

copyright-year

2015

license

license-p

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 3.0 License.

abstract

title

Abstract

This study describes the influence of the phosphorothioate internucleotide bond on the deoxyribonucleic acid (DNA) oxidation process. The interaction of an ultraviolet radiation (UVA) with a targeted double-stranded (ds) oligonucleotide, in which one strand contains an anthraquinone (AQ) moiety on the 5'-end, may lead to a hole migration process through the double helix. In the end, the migration of the formed radical cation terminates in a suitable place. Usually, this is a guanine-rich

Error

Warning

RNG

-

 RNG_sc-set 3

A ▾

business-rules

-

 journalid-onlinecode 3

B ▾

 journalid-doicode 3

C ▾

 unknown-value 7

D ▾

 journal-code-online 3

E ▾

 journal-title-unknown 3

F ▾

 issue-article-count 1

G ▾

 article-pub-date-epub 2

H ▾

 date_day 1

I ▾

 date_month 2

J ▾

 mandatory-element 1

K ▾

 package-basename 1

L ▾

 folder-structure 16

M ▾

 match-article-id-doi 3

N ▾

 article-id-regex 3

O ▾

 doctype-public 3

P ▾

Integrating Schematron in the Development Process



Validate XML with Schematron

- Associate Schematron in the XML file

```
<?xml-model href="books.sch" type="application/xml"
             schematypens="http://purl.oclc.org/dsdl/schematron"?>
```

- Use tool-specific association options
 - Associate Schematron file with a set of XML files (all files with a specific namespace, or from a directory)
 - Associate Schematron with all XML files from a project

Run Schematron Validation

- From an XML editing framework
 - Check the XML files as you type
 - Run the Schematron validation on multiple files
 - Generate a report

Run Schematron Validation

- Using W3C's XProc pipeline language through its "validate-with-schematron" step

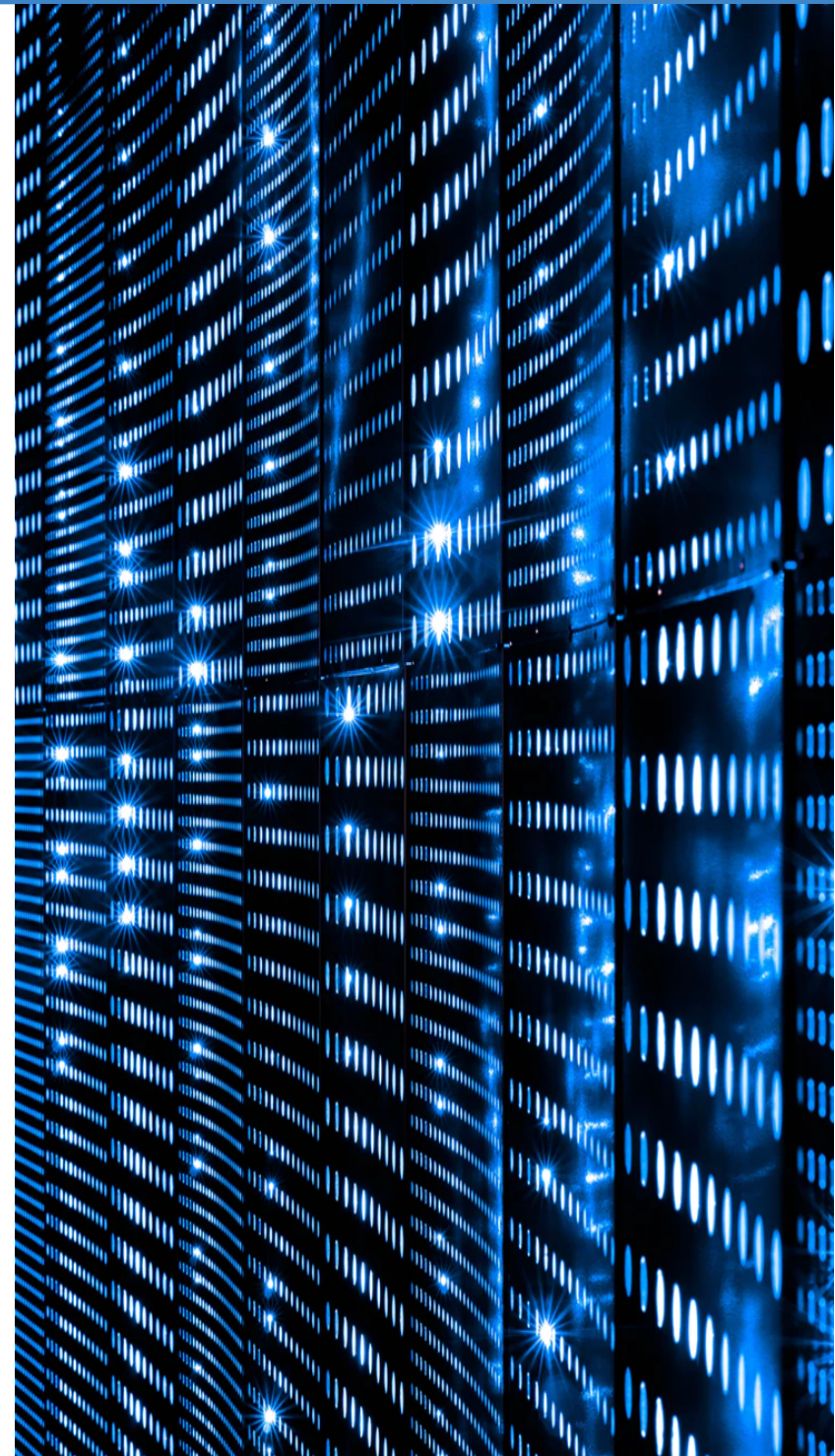


- Using Apache Ant, from bat/shell
<https://github.com/Schematron/ant-schematron>

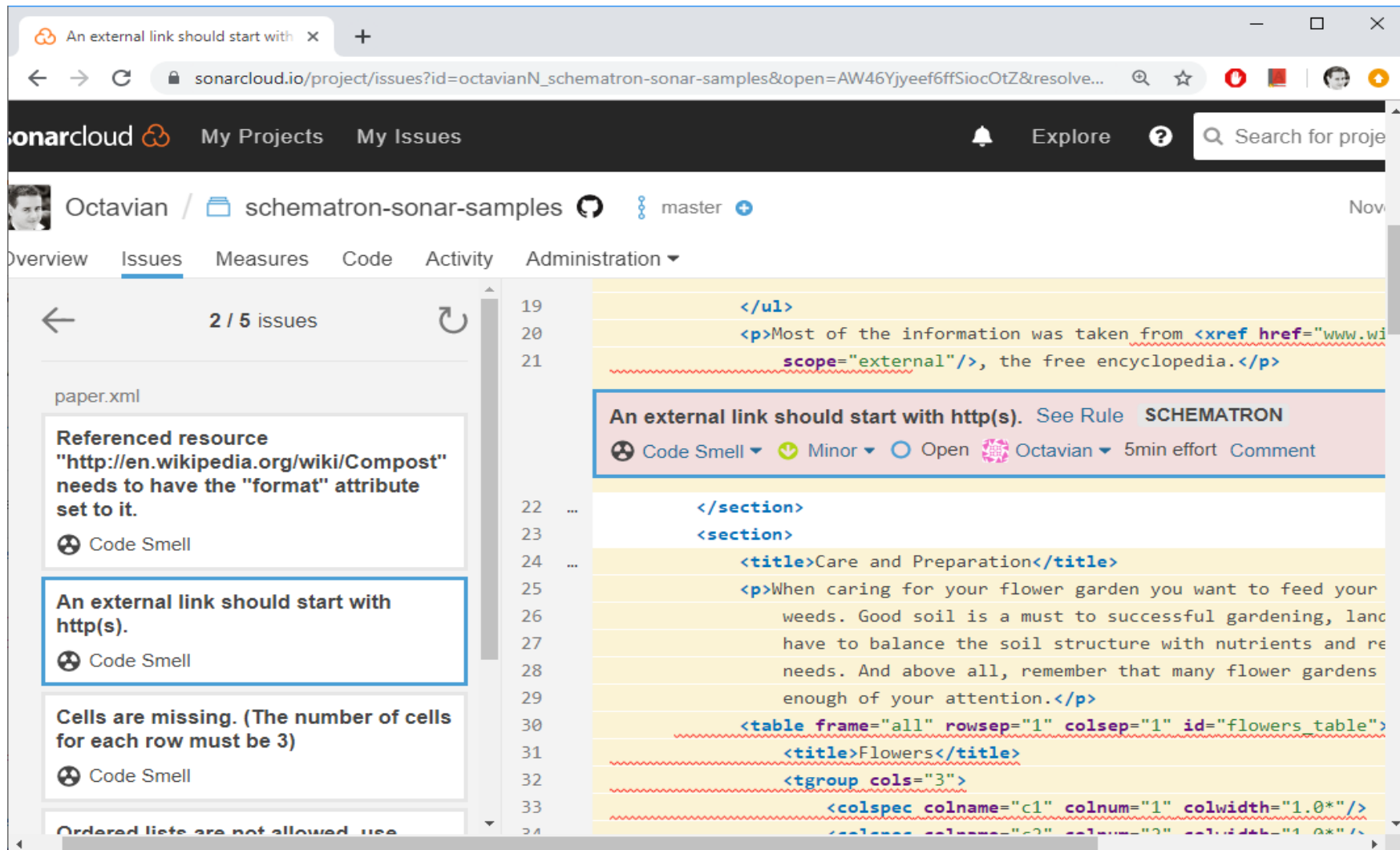


Integration server

- Use an integration server to run Schematron checks (such as Jenkins, Travis)
- Send a report to the user



Sonar Schematron



The screenshot displays the SonarCloud web interface for a project named 'schematron-sonar-samples'. The 'Issues' tab is active, showing a list of 2 out of 5 issues. The selected issue is a 'Code Smell' with a severity of 'Minor', titled 'An external link should start with http(s)'. The issue is associated with the rule 'SCHEMATRON' and is located in the file 'paper.xml'. The code editor shows the XML content with the issue highlighted in a pink box. The issue description in the left sidebar is also highlighted in a blue box.

2 / 5 issues

paper.xml

Referenced resource
"http://en.wikipedia.org/wiki/Compost" needs to have the "format" attribute set to it.
Code Smell

An external link should start with http(s).
Code Smell

Cells are missing. (The number of cells for each row must be 3)
Code Smell

Ordered lists are not allowed, use

```
19     </ul>
20     <p>Most of the information was taken from <xref href="http://en.wikipedia.org/wiki/Compost"
21         scope="external"/>, the free encyclopedia.</p>
22 ...
23     <section>
24 ...
25     <title>Care and Preparation</title>
26     <p>When caring for your flower garden you want to feed your
27         weeds. Good soil is a must to successful gardening, land
28         have to balance the soil structure with nutrients and re
29         needs. And above all, remember that many flower gardens
30         enough of your attention.</p>
31     <table frame="all" rowsep="1" colsep="1" id="flowers_table">
32         <title>Flowers</title>
33         <tgroup cols="3">
34             <colspec colname="c1" colnum="1" colwidth="1.0*" />
```

An external link should start with http(s). See Rule SCHEMATRON
Code Smell Minor Open Octavian 5min effort Comment

Conclusion



- Schematron can be used for documentation QA
- Problems with different severity
- Notification about the problem as you type
- Generate a QA report





Questions?

Octavian Nadolu
Software Architect at Syncro Soft

octavian.nadolu@oxygenxml.com

Twitter: [@OctavianNadolu](https://twitter.com/OctavianNadolu)

LinkedIn: [octaviannadolu](https://www.linkedin.com/in/octaviannadolu)

