

Schematron

Andrew Sales

Andrew Sales Digital Publishing

XML London, 11th June 2017

About

- XML validation beyond schemas since 2000
- 2016: Project Editor, ISO/IEC 19757-3 (Schematron)

WARNING: personal opinions/evangelism follow!

What is Schematron?

What is Schematron?

- “Sleeper” standard
- No great PR or fanfare; now quite widely used
- Simple idea; just works

“May well be the XML project’s greatest technical legacy to the world.”

- Simon St Laurent

What is Schematron?

“A feather duster to reach the corners that other schema languages cannot reach.”

- Rick Jelliffe, inventor of Schematron

- Means of expressing business rules
- Not grammar-based
- Greater flexibility to check XML
- Just an XML format: XPath welcome

Excursus: before Schematron

- Bespoke validation (SAX handler, XSLT...)
- In-house conventions
- Maintenance costs

*Would you write your own
schema/transformation language?*

Examples

- Table checking
- Co-occurrence constraints
- Remote resources
- *etc etc*

Benefits (implementers)

- Approach to expressing constraints formalized
- No need to roll your own
- If it can be computed, Schematron can check it
- Inclusions -> modularity
- Abstraction -> code re-use, lower maintenance

Benefits (managers)

- Open standard
- Main implementation free, open source
- Good tool support
- Ease of maintenance
- Reduced reliance on bespoke tools & their keepers...
- Better, faster feedback loop to improve content quality

Benefits (users)

- Separation of duties:
 - Person A implements the rule, writes the code
 - Person B writes the message and how to fix it
- Experts concentrate on their own area of expertise

Validity redefined

- No longer a simple binary condition
- Not just “is it invalid?”, but “**how** is it invalid?”
- Fine-grained measure of validity
- Validation results are *themselves* XML, so can be stored, queried, processed...
- Metrics to gauge validity e.g. over time
- Opens way to **content scoring**

“Agile” validation

- Flexibility to tailor:
 - validation to stage in workflow
 - messaging to audience

Best practice

- Treat like any other software:
 - documentation
 - test suites
- Tie schema constraints to business rules early
 - rationales are easy to forget

There are no quick fixes(?)

- Schematron QuickFix
- Extend your schema with foreign elements
- Automated fixes

Resources

- ISO standard: <http://bit.ly/2kw5FxK>
- <https://github.com/Schematron> (thank you, Tony!)
- schematron.com
- schematron-quickfix.com
- [nigelwhitaker/cals-table-schematron](https://github.com/nigelwhitaker/cals-table-schematron)
- GitHub: [AndrewSales/XMLLondon](https://github.com/AndrewSales/XMLLondon)