



RUHR-UNIVERSITÄT BOCHUM

# Bridging the Gap: Secure and lossless conversion of XML data structures to the JSON format

Bachelor thesis ■ March 30, 2017 – June 29, 2017

**May 09, 2017**

**Advisors: Dennis Felsch & Paul Rösler**

Jan Holthuis

- 1 Quick Recap
- 2 Progress report
- 3 Conversion
- 4 Next steps

# Last time we learned that...

- XML parsers can be vulnerable to attacks from these areas:
  - Denial of Service (Billion Laughs, Quadratic Blowup)
  - File System Access (XXE, XInclude, ...)
  - Server-Side Request Forgery (DTD Retrieval, Schema Location, etc.)
- I implemented a framework for automatic testing and added security testcases for these vulnerabilities

- 1 Quick Recap
- 2 Progress report
- 3 Conversion
- 4 Next steps

# Progress made in the last two weeks

xjcc Tool Development

- Added some minor improvements on checking tool and some more security testcases
- Spent a lot of time reading W3C recommendations
- Started establishing conversion quality criteria

- 1 Quick Recap
- 2 Progress report
- 3 Conversion**
- 4 Next steps

- Make sure that document content and logical structure stays the same...
- ...but the physical structure is not important.
- Thus we can use the XPath Data Model and Canonicalization for abstraction

# Conversion criteria examples

## Comments

*They are not part of the document's character data; an XML processor may, but need not, make it possible for an application to retrieve the text of comments. (Extensible Markup Language (XML) 1.0 (Fifth Edition), W3C, 26 November 2008, Section 2.5])*



# Conversion criteria examples

## Document Order: Elements

*There is an ordering, document order, defined on all the nodes in the document corresponding to the order in which the first character of the XML representation of each node occurs in the XML representation of the document after expansion of general entities. Thus, the root node will be the first node. Element nodes occur before their children. Thus, document order orders element nodes in order of the occurrence of their start-tag in the XML (after expansion of entities). (XML Path Language 1.0, W3C, 19 November 1999, Section 5)]*

# Conversion criteria examples

Document Order: Attributes

*Note that the order of attribute specifications in a start-tag or empty-element tag is not significant. (Extensible Markup Language (XML) 1.0 (Fifth Edition), W3C, 26 November 2008, Section 2.5]*

*The relative order of namespace nodes is implementation-dependent. The relative order of attribute nodes is implementation-dependent. (XML Path Language 1.0, W3C, 19 November 1999, Section 5]*

- 1 Quick Recap
- 2 Progress report
- 3 Conversion
- 4 Next steps**

- Research and establish more conversion criteria and add test docs
- Evaluate current solutions using the test documents
- Possibly develop custom algorithm

# Questions?

Reach out via email:

- **Jan Holthuis**  
[jan.holthuis@rub.de](mailto:jan.holthuis@rub.de)

