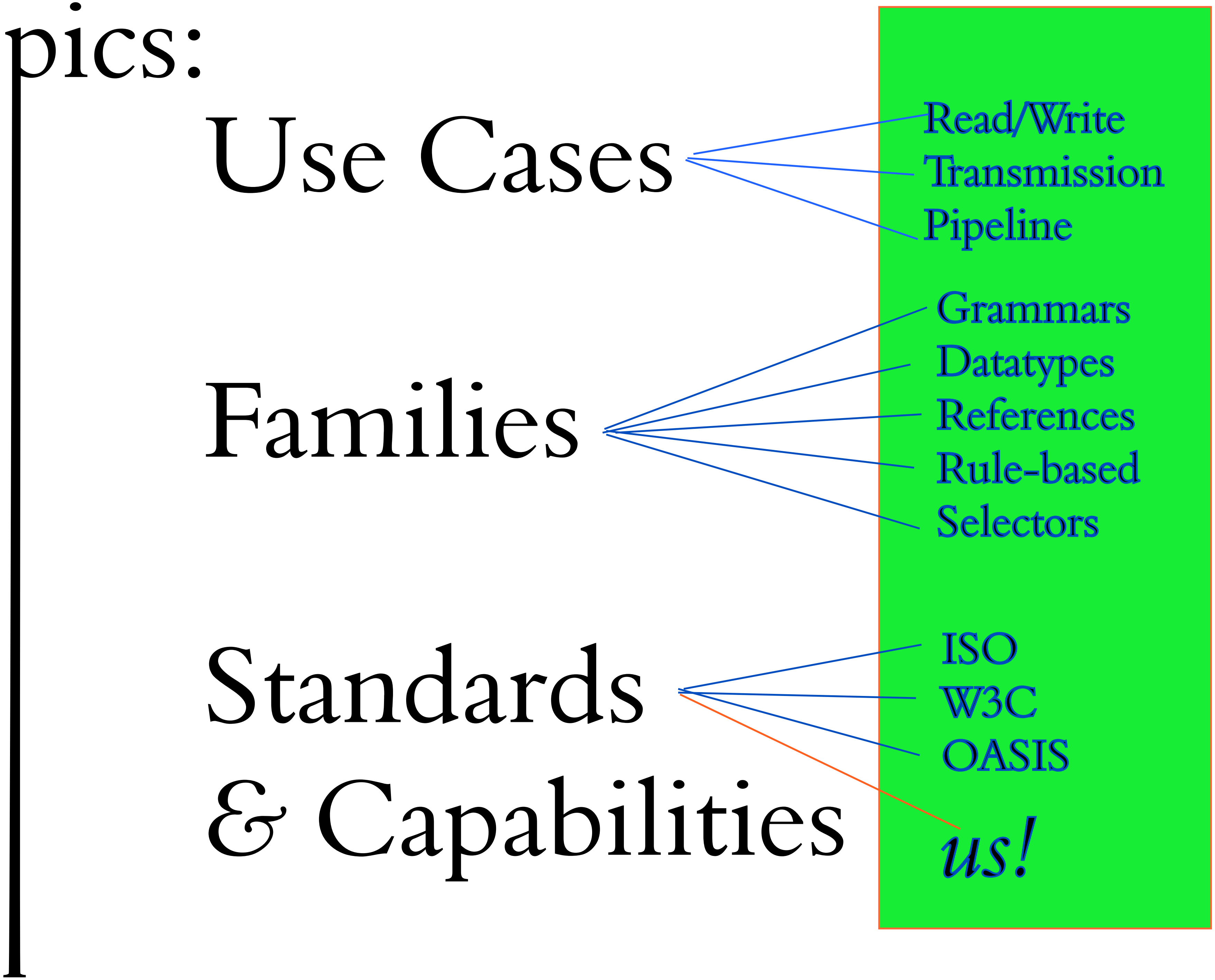
Document Validation

Rick Jelliffe
Topologi, Pty. Ltd.
ricko@topologi.com

Topics:



Use Cases

- Read/Write
- Transmission
- Pipeline

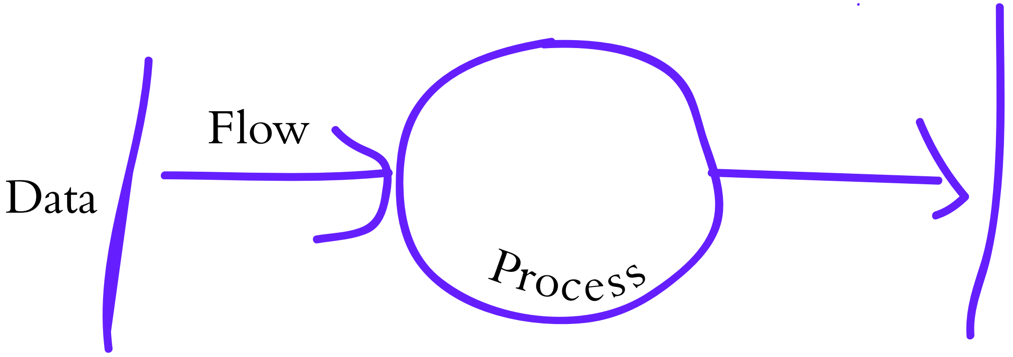
Families

- Grammars
- Datatypes
- References
- Rule-based
- Selectors

Standards & Capabilities

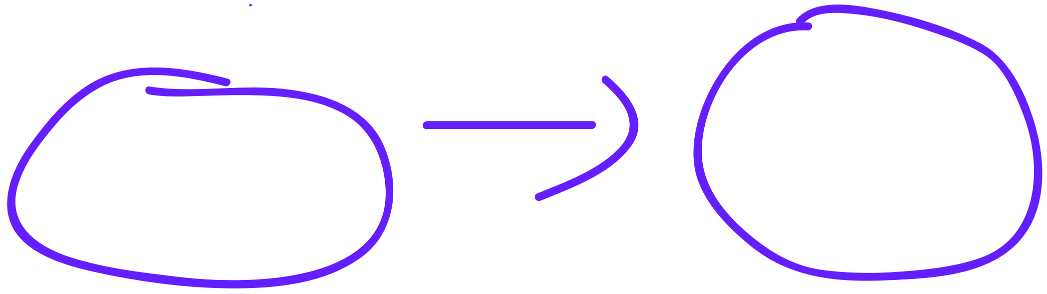
- ISO
- W3C
- OASIS
- *us!*

Use Cases for XML



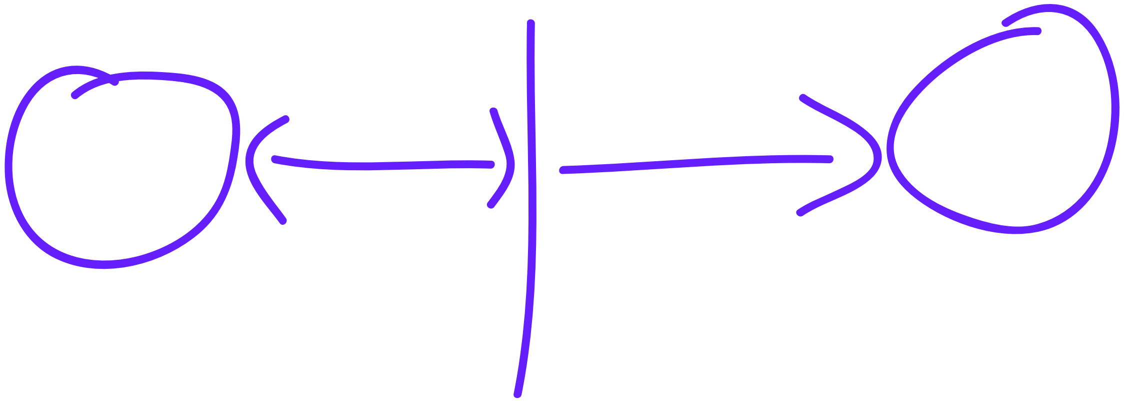
Transform

? Web Serving



Pipeline

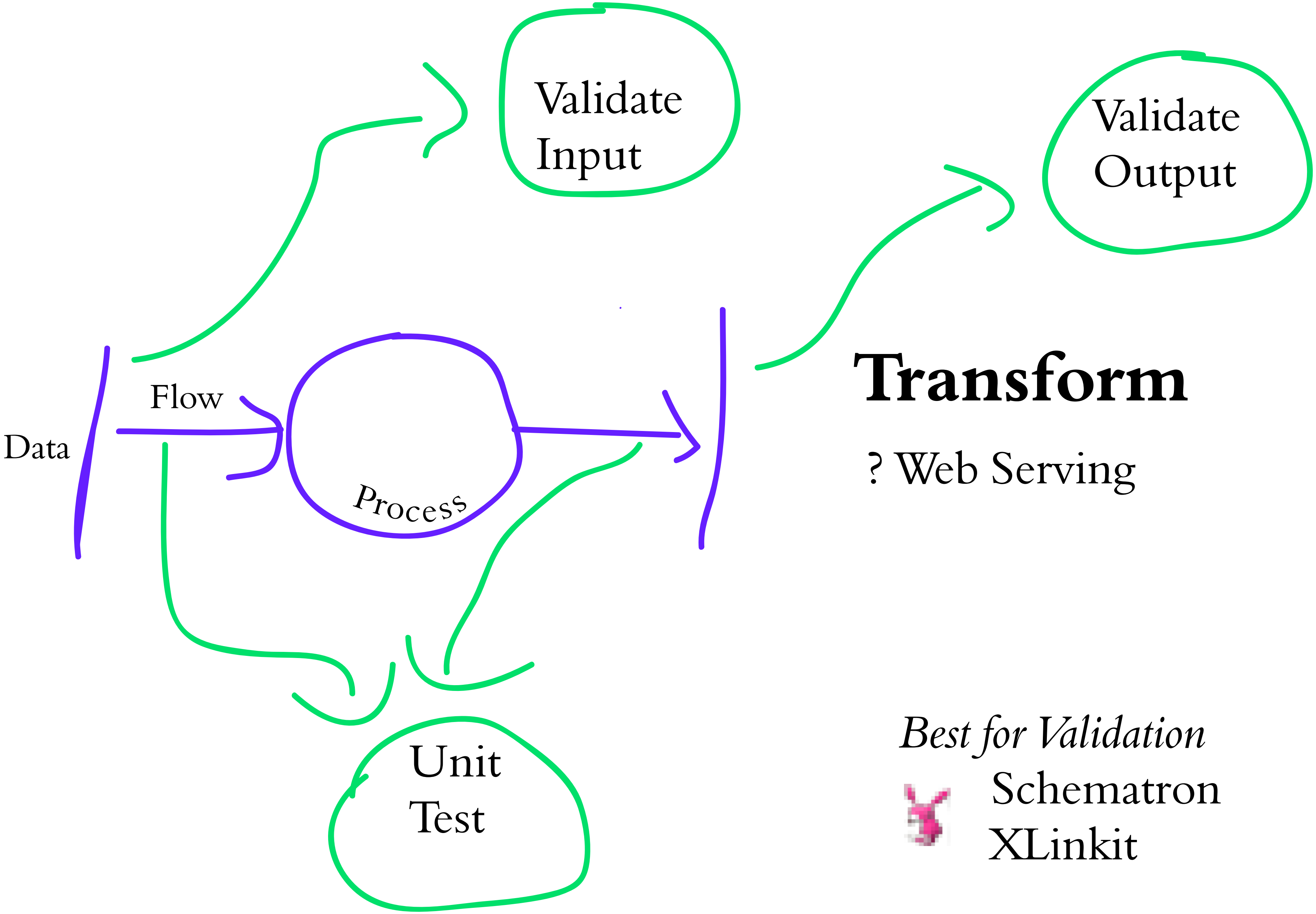
? Data Services



Read/Write

? Content Construction

Use Case



Validate Input

Validate Output

Data

Flow

Process

Unit Test

Transform

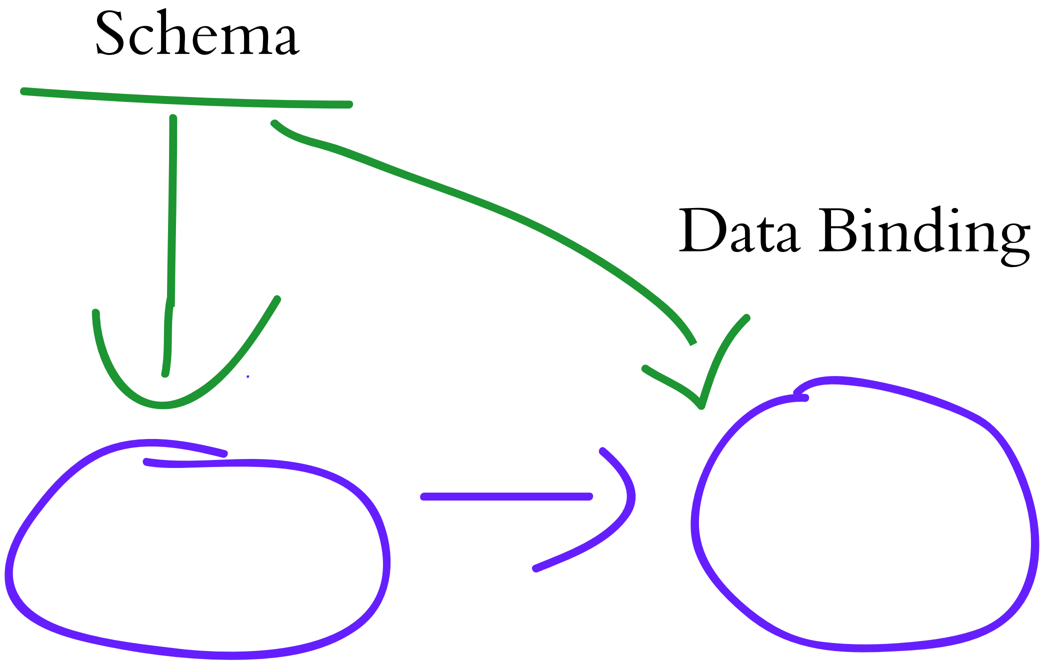
? Web Serving

Best for Validation

Schematron

XLinkit

Use Case



Schema

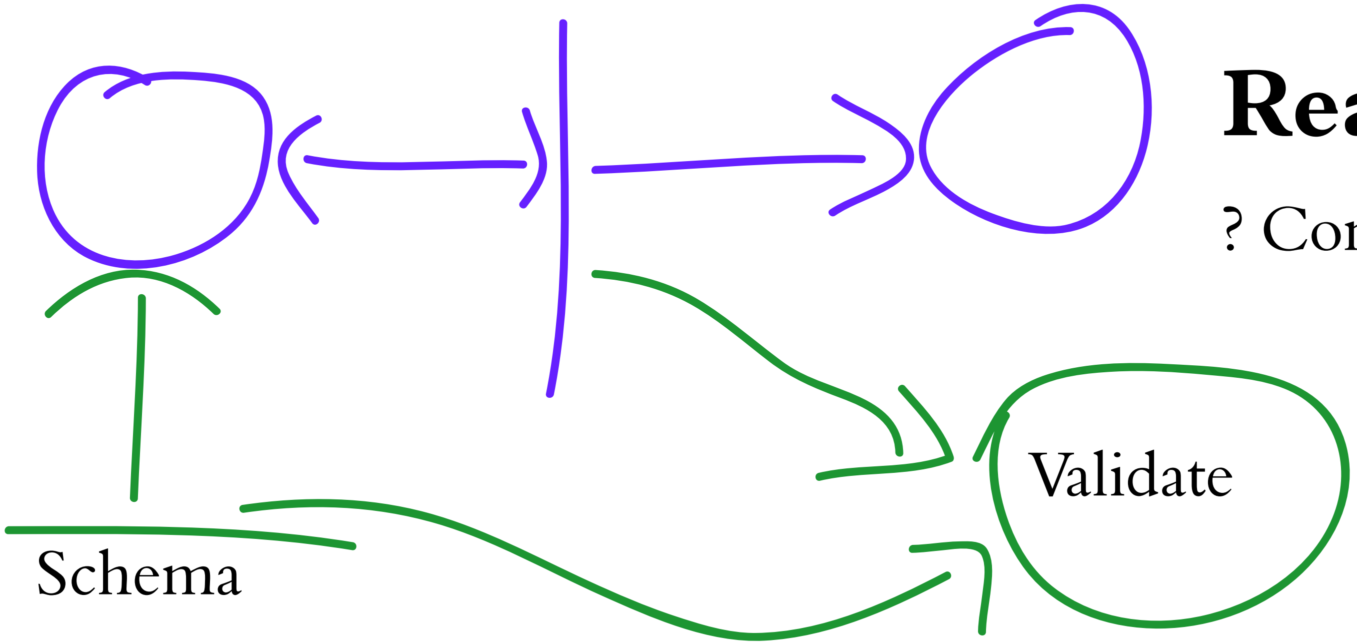
Data Binding

Pipeline

? Data Services

Best for Data Binding
W3C XML Schema

Use Case



Read/Write

? Content Construction

Validate

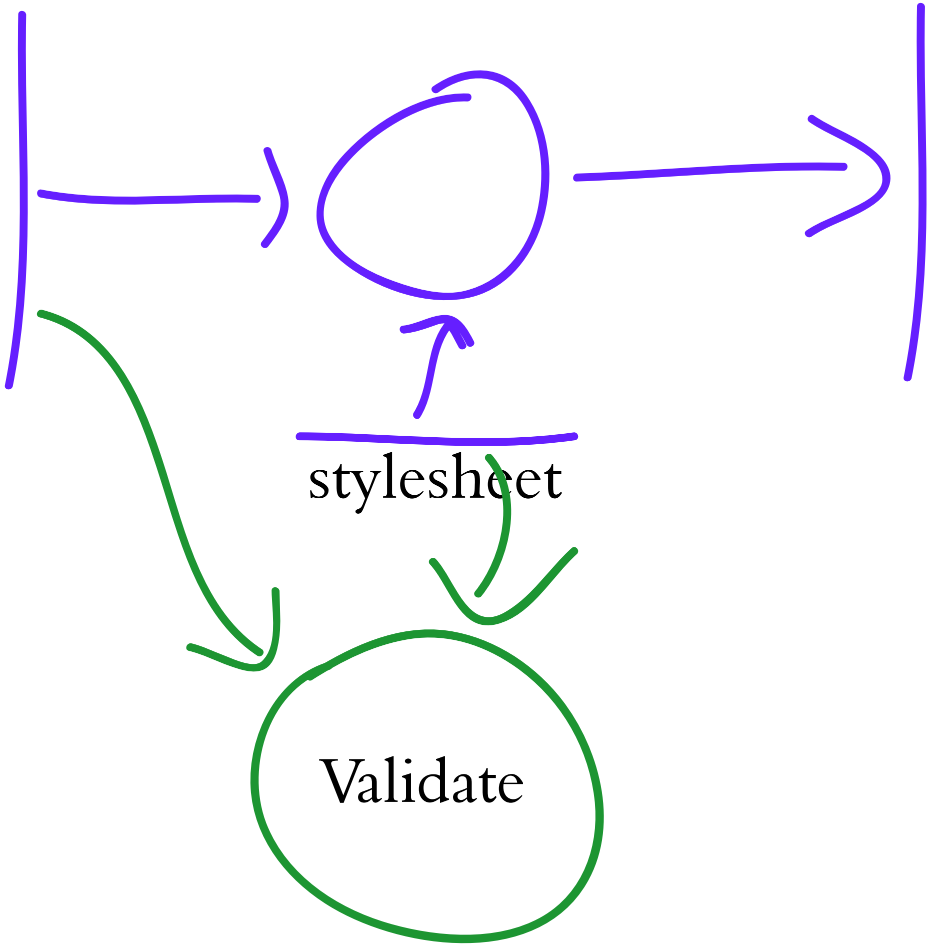
Schema

Best for Publishing

RELAX NGconvert from DTD (e

convert to WXS for editors

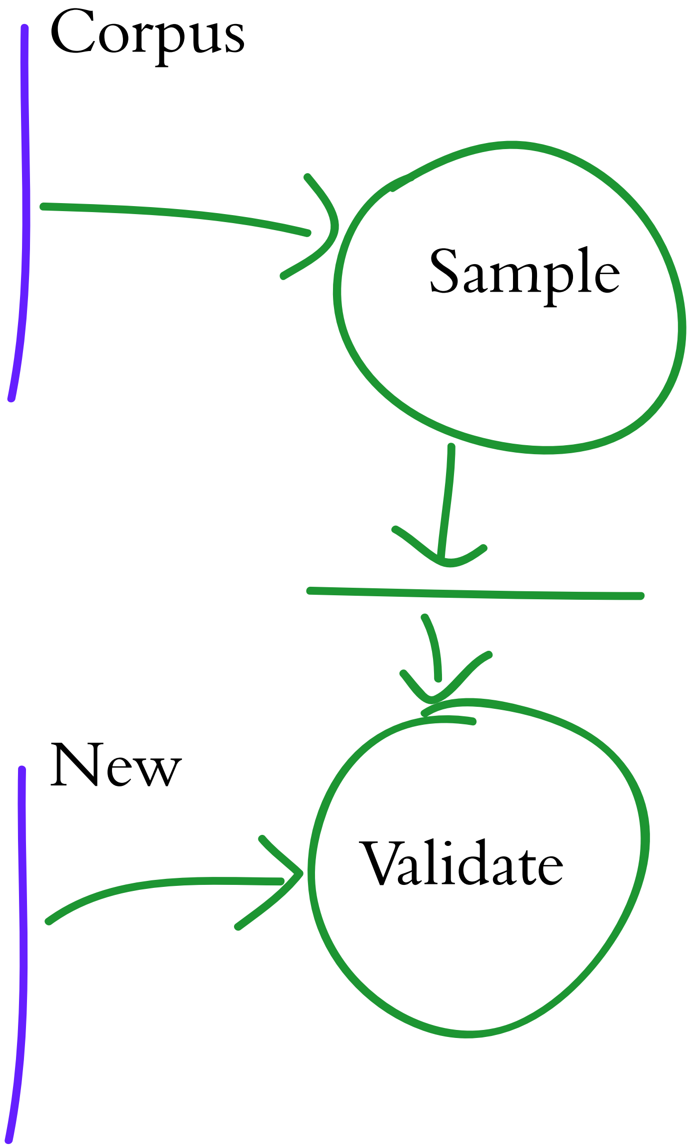
Use Case: New Field



Validate Documents against Stylesheets

Example
3B2

Use Case: New Field



Corpus

Sample

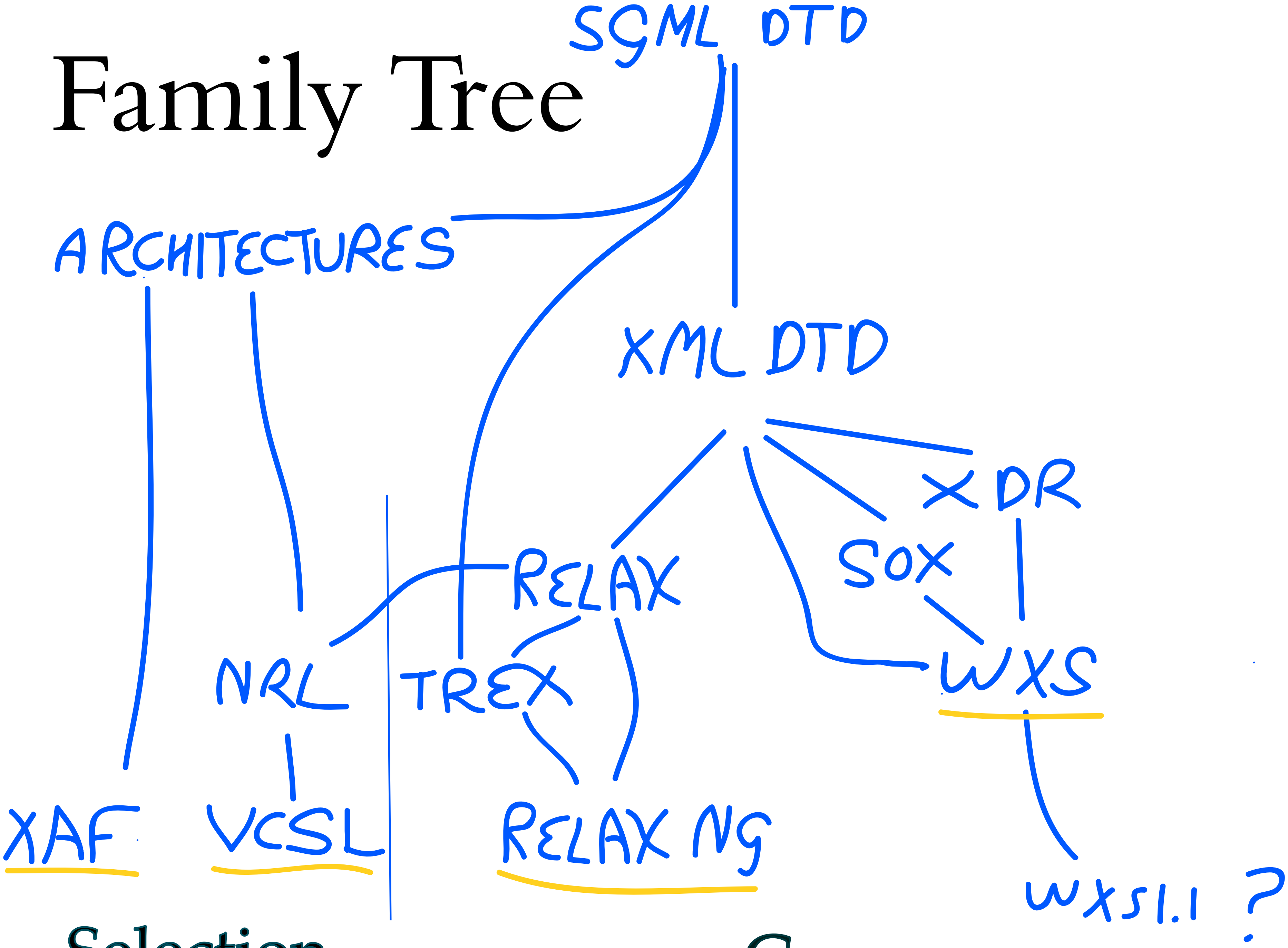
New

Validate

Validate Documents against Samples

Example
Topologi Reporter

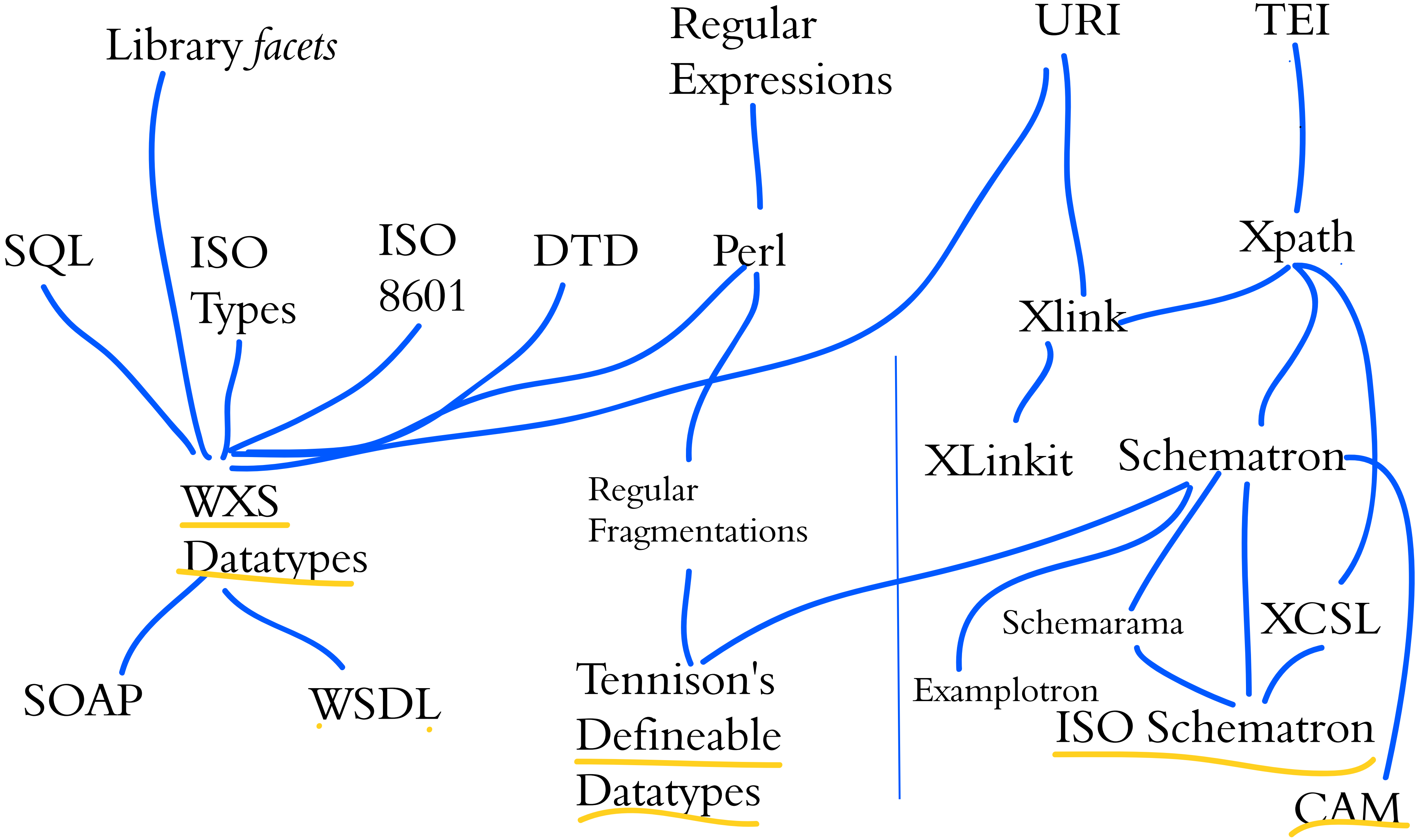
Family Tree



Selection

Grammars

Family Tree



Library *facets*

Regular Expressions

URI

TEI

SQL

ISO Types

ISO 8601

DTD

Perl

Xpath

Xlink

WXS Datatypes

Regular Fragmentations

XLinkit

Schematron

SOAP

WSDL

Tennison's Defineable Datatypes

Schemarama

XCSL

Examplotron

ISO Schematron

CAM

Datatypes

Rule-Based

Seedlings

Referential Integrity

╰─WXS, Schematron, ...?

Character Classes

╰─Character Collections, ...?

Example Based

╰─Examplotron, ...?

Partial Order-based

╰─Hook, ...?

Logic-based

╰─?

Model-based

╰─CAM, UML, ... ?

+ Namespace-aware DTDs?

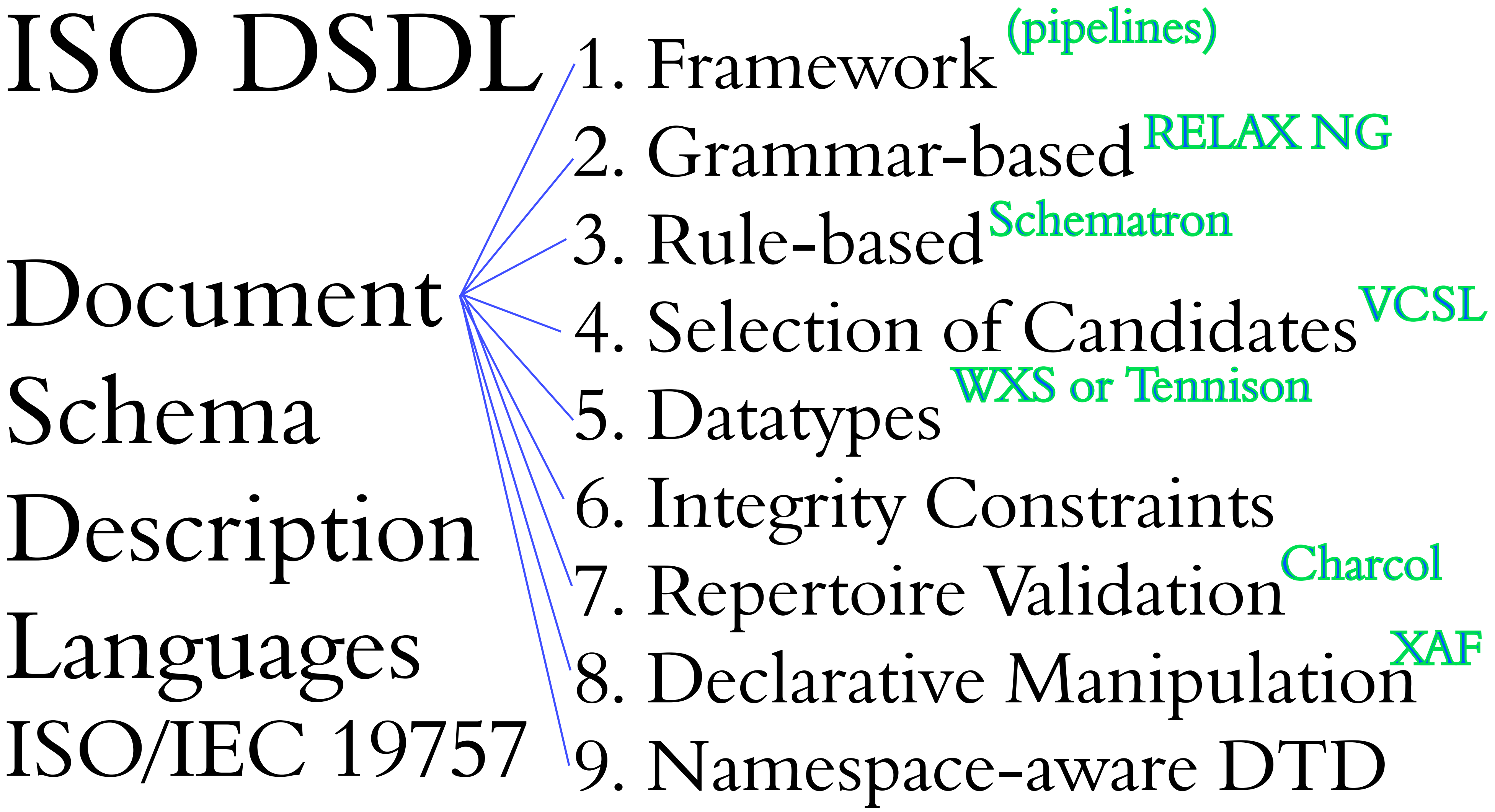
Exciting!

More activity, experiments and innovation than ever before.

Simple schemas can use any language

W3C XML Schema helps great innovation for data uses: little for document uses

How can industrial publishing make use of these innovations ?



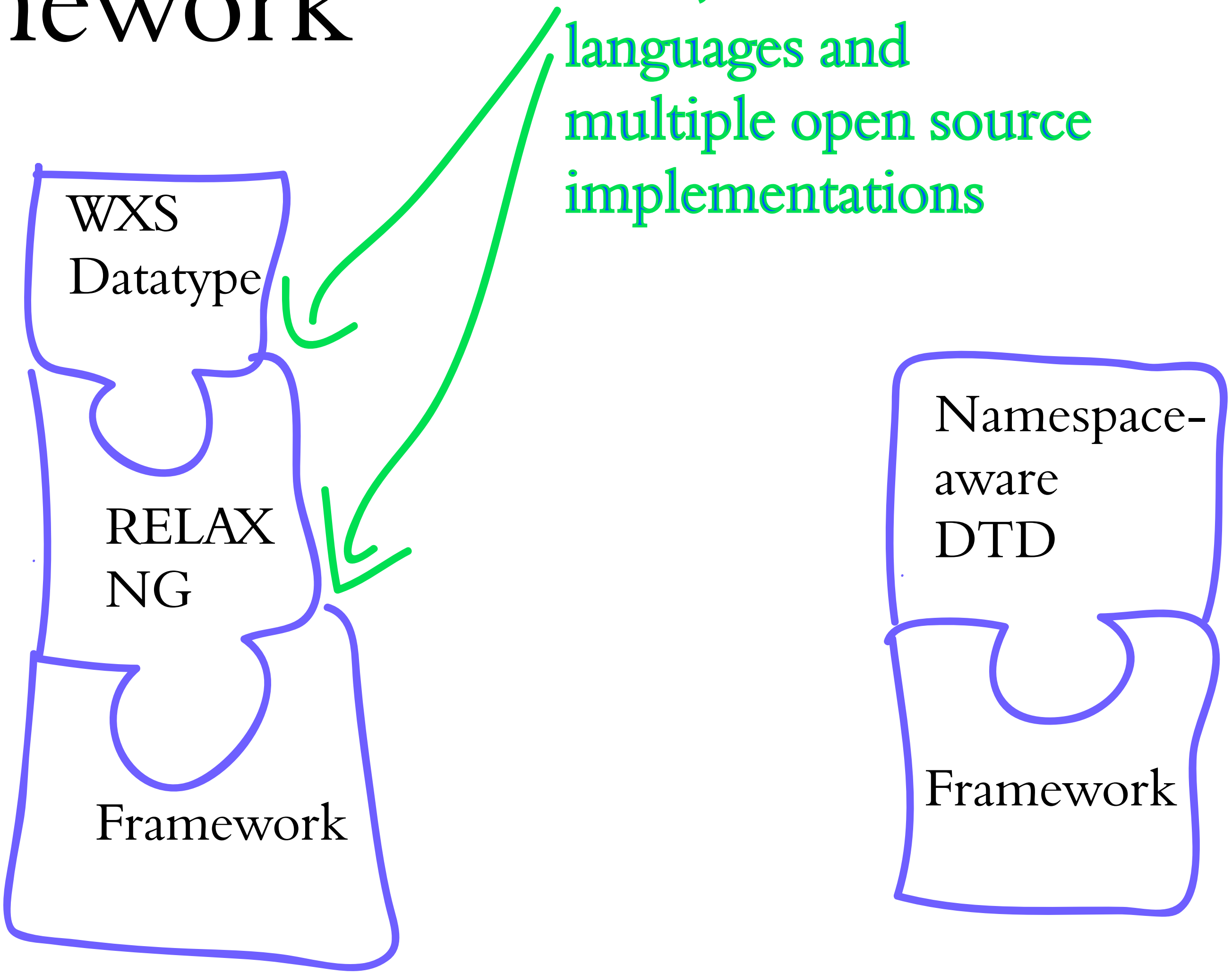
ISO DSDL

Document
Schema
Description
Languages
ISO/IEC 19757

1. Framework (pipelines)
2. Grammar-based RELAX NG
3. Rule-based Schematron
4. Selection of Candidates VCSL
5. Datatypes WXS or Tennison
6. Integrity Constraints
7. Repertoire Validation Charcol
8. Declarative Manipulation XAF
9. Namespace-aware DTD

ISO DSDL

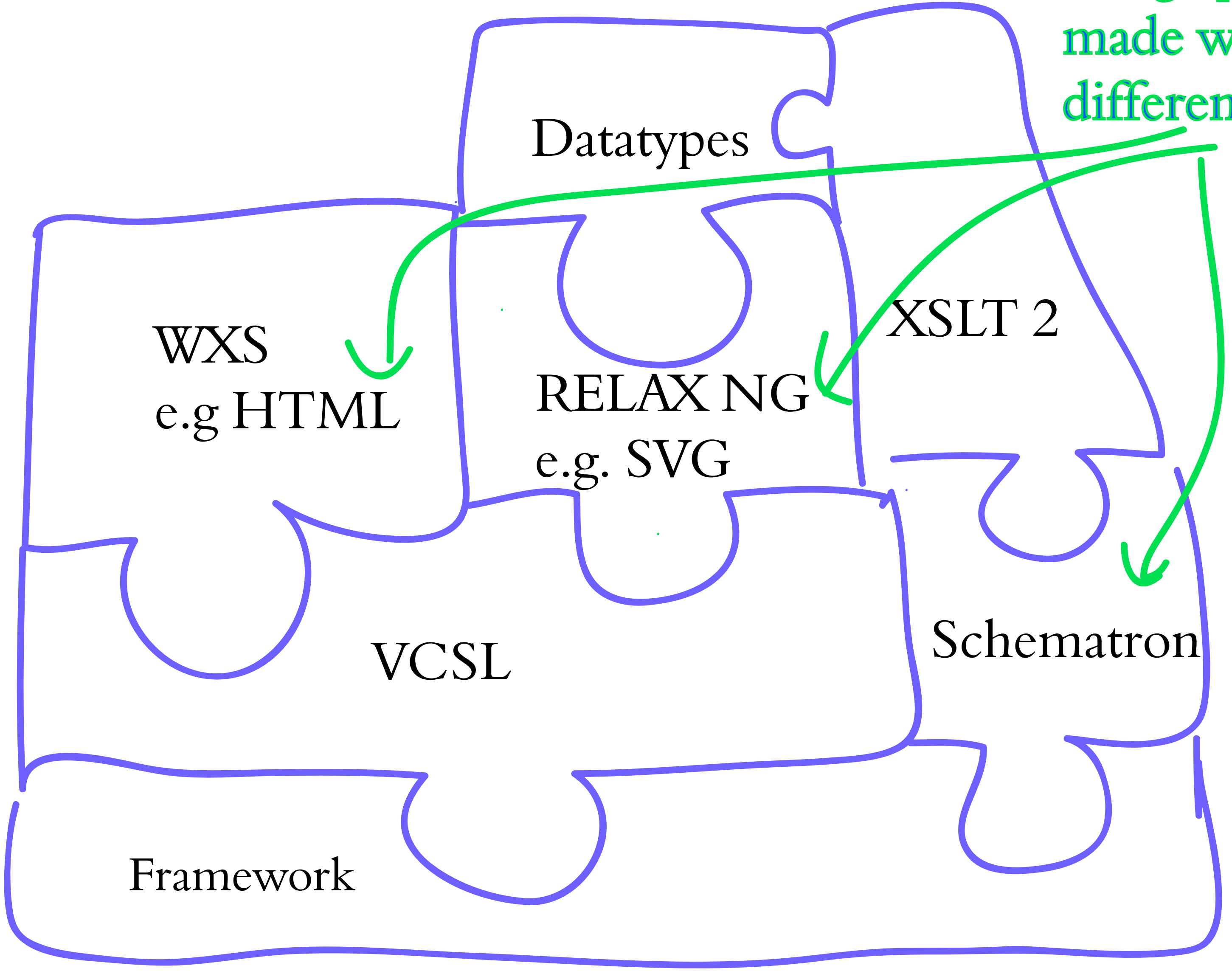
Framework



Plug-in Modularity encourages small, learnable languages and multiple open source implementations

ISO DSDL

Plug-in Modularity allows you to merge projects made with different schemas



Datatypes

WXS
e.g HTML

RELAX NG
e.g. SVG

XSLT 2

VCSL

Schematron

Framework

VCSL

Validation Candidate Selection Language

from Murata's RELAX Namespace + Clark's NMS and NRL

1 Select a branch using namespace and context

2 Trim out elements you don't want

3 Send this to one or more validators

Solves *big* problem with namespaces: they promise modularity, but you need monolithic schemas.

E.g XHTML+RDF+SVG+DC+MathML

Schematron

Rule-based Validation

```
<rule context="series">
 <let name="first"
        value="string-before(., '-')" />
 <assert test="number( $first ) =6">
  The value of the first part must = 6
  </assert>
</rule>
```

```
<series>5-32-86</series>
```

News:

- <let> allows variables
- Abstract patterns you to make your own schema language
- Specified as framework, allows migration to XSLT2
- <key> cleaned up
- Draft specification out this week